

```

AC 09R017;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3.
GN ZAP3 OR ZAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RA Misawa K., Nosaka T., Kitamura T.;
RT "A huge nuclear protein rich in proline similar to human hypothetical
RT protein zap3 and zap13."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: Nuclear (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB03168; BAA85182.1; .
DR MGD; MGI:1926195; Zap3.
KW Nuclear protein.
FT DOMAIN 15 204 PRO-RICH.
FT DOMAIN 355 473 GLN-RICH.
FT DOMAIN 925 1012 ARG-RICH.
SQ SEQUENCE 1386 AA; 155130 MW; D862F918ED221DF CRC64;

Query Match 5.68; Score 181.5; DB 1; Length 1386;
Best Local Similarity 20.98; Pred. No. 0.3; 266; Indels 237; Gaps 30;
Matches 149; Conservative 61; Mismatches

QY 34 PEDPEPAEAHFLRRRRKLEKPPPLASSGSGSPAPSPACGGKGLLPAGAPGQOE 93
DB 148 PESPPVPGSYMPSPSYMPPOPPPSYVPPSSAQP-----YLPAPPSPSKFO 196
QY 94 ESMGGSVLPICPP--PATKQAGIGGEPAAGA-----GCSPPR----- 129
DB 197 -----LPPPSIPSGNKTALIQEPLESGAKNKTAEQKQAPDPDPSTMTPEQQQYW 248
QY 130 -----KYAVLPIQTGSLVAAKERTPMAGDKGGAAPATAASDAPGPPPLP- 178
DB 249 YRQHLISLQORTKVHLPGHKKGGLVTAKDVPEPIKEAPGPAASQVAEPIAAEKPPPLPPPN 308
QY 179 --GPPPLAP-----TATAGTTLAASGRMKSMKSPSLGGGGS 213
DB 309 EEAAPPPLSPREPSSEDSSEDESEDAEFKQLKALAHQAALAHQOQOQOQYV-GEYOY 367
QY 214 GASSQAACLKQILLDLLEQOQOOLQAKKEIEELKESRDTLARIERMERMOLVKK 273
DB 368 GIMQRTTLOQLI-----QQYQOVIQHSPIQTM--SLDVQLRHYEQMOOQOFLFQ 417
QY 274 DNEKE-----RHK-LFGVYTE-----ERETELSEXT-----K 301
DB 418 DWEREQVLEBQHLSTPHKQOLOEVEQKWTQOGHMKATQYLOEKVNSFQTVKSOYLG 477
QY 302 LECQPELSETQTLPPKPFSCGRSGKHKRKSFGSTERTKTPVKIAPESKVKTKPKH 361
DB 478 MAMPPEFVPSQMPPLP-----TMPPVPLP--SLPPVAVPRA 514
QY 362 SPKEPCGSLSTVCRELKSOETPEKPRS-----SDTPRELSTPQKGPSTHREKA 415
DB 515 LPTSIPTPPGMPVPVMPPLPTSVPPGMPPLSLSSGPPVLPPLPALSAGSPVLPALP 574
QY 416 FSSEIDLPYLTSTTEMYLCRMHQPP-----SPLPLRESPPKE--ETV 457

```

```

DB 575 GCPPIILPLPLSSA-----TPPGIIPPGAPAGQMPQTLTAPLPASSGQNSQIPEKP 626
QY 458 ARCLMPSVA-GFTSVLAVPSMRDHSVEPIRDNPDLLENLD-DVEFSRHAHKLLEDEK 515
DB 627 RQALLPLPVSEGSTP-----PS-----PYHPPQSEQVNSKPLNKFSSQD--LGES 672
QY 516 RRRKWDIORIREORIL-----ORLOLRMYKKKGIOSEPEVYS 553
DB 673 SALSQSLIAKQDPVPSGGLLADPPKGSFLEGRRGREGKEQJQKLDKG---SEQMD 729
QY 554 FPEPDDVESLMTTP-----FLPVAFGRP-----LPKLPQNFELP 550
DB 730 HLPPPD-----SLQNPSPRGWYPPPGSVRRPPPMGKPPGSIVRSPAPASSIP 779

RESULT 11
SHK1_RAT
ID SHK1_RAT STANDARD; PRT; 2167 AA.
AC Q9WV48; Q9WU13; Q9WUE8; Q9QZ48;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE S13 and multiple ankyrin repeat domain protein 1 (Shank1) (GKAP/SAPAP
DE interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
DE interacting protein) (SSSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RP DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Tao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein."
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRATIN-Sprague-Dawley;
RX MEDLINE=99360550; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Morley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin."
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor C11 interacts directly with proteins
RT of the Shank family."
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RP DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
RT genes, alternative splicing, and differential expression in brain and
RT development."
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoernck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
RT multidomain proteins present in human and rodent brain."

```

GN SHANK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1: 2 AND 3), AND INTERACTION WITH SSTR2.
RC TISSUE=Fetal brain, Hippocampus, and Thalamus;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenc H.-H., Baechner D., Richter D., Krielenkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [2]
RP REVIEW.
RX PubMed=10806096;
RA Sheng M., Kim E.;
RT "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
density (PSD) of excitatory synapses that interconnects receptors
of the postsynaptic membrane including NMDA-type and metabotropic
glutamate receptors via complexes with GKAP/PSD-95 and Homer,
respectively, and the actin-based cytoskeleton. May play a role in
the structural and functional organization of the dendritic spine
and synaptic junction.
CC -1- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with
SPAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2 (By similarity).
Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
neuronal cells (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1/a (shown here), 2/b and 3; are
produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in brain particularly in the
amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
seems to be expressed ubiquitously.
CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DR EMBL; AF163302; AADA5121.1; -
DR EMBL; AF226728; AAF35887.1; -
DR HSSP; P06241; 1SHF.
DR MIM; 604999; -
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001452; SH3.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.

DR PROSITE; PS50002; SH3; 1.
KW ANK Repeat; SH3 domain; Repeat; Alternative splicing.
FT REPEAT 212 245 ANK 1.
FT REPEAT 246 278 ANK 2.
FT REPEAT 279 312 ANK 3.
FT REPEAT 313 345 ANK 4.
FT REPEAT 346 378 ANK 5.
FT REPEAT 379 395 ANK 6.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2098 2161 SAM.
FT DOMAIN 1002 1007 POLY-HIS.
FT DOMAIN 1014 1019 POLY-HIS.
FT DOMAIN 1189 1195 POLY-GLY.
FT DOMAIN 1709 1717 POLY-GLY.
FT DOMAIN 1844 1854 POLY-PRO.
FT DOMAIN 1896 1902 POLY-GLY.
FT DOMAIN 1970 1979 POLY-SER.
FT VARSPIC 1 613 MISSING (IN ISOFORM 2).
FT VARSPIC 614 654 MISSING (IN ISOFORM 3).
FT VARSPIC 646 654 MISSING (IN ISOFORM 3).
FT SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;
SQ
Query Match Best Local Similarity 22.5%; Score 182.5; DB 1; Length 2161;
Matches 128; Conservative 44; Mismatches 191; Indels 207; Gaps 28;
OY 32 GGPDEPCGAHPLPRHRLKEPPPLASSOGGSPASP-----ACGGGKRG 80
DB 1030 GSPDDP-----PRLALGPQPSLRGRRGGPPTPAPSPSHHSGAGGGGSSQG 1079
OY 81 -----LLLPAGAA-----PCQEESSMGVPLPCPPATQAGIGEPAAAGACSPRP 129
DB 1080 PALKTIQPLPPAASAMVVPARGRGK-----PLVQTKEGEP-QKGGGLPPAP 1130
OY 130 KYQAVLPLOTGLVAAKE-----PTPWAG 154
DB 1131 SPSSPASQPPPAVAAPSEKNSIPITIIKAPSTRSSGSSGSTEAEPTQDEPTGG 1190
OY 155 DKGAASPA-ATASDPAGPPLPLP-----GPPLIATYATAGT-----LAASEGRW-KS 201
DB 1191 GGGGSSPSPAPAMSPVPPSPVPTPASPGPATLDTSGFAGALVGAAAREGQWNEA 1250
OY 202 MKRSPL-----GGGSGASQAACLKQILLDLDEQQOOLQAKEREIEELK 251
DB 1251 RRRSTLPLSTAGDEDDGDLGTGA-----APGRRLRISK 1286
OY 252 SERDTLLARIERMERRMQLVKKDNEKERHKLFGYTEBERETELSEKIKLECOPELSET 311
DB 1287 SIDEGMFS-----AEYHLRLAESAGSGAG-----YGGVGAASRAYVGGGGS-SAF 1329
OY 312 SQTLPKPPSCGRSGKGRKRSFPG-----STER-----KTVKKLAEFEFSVK 355
DB 1330 TSFLPPRPVLVPLGTGKALDPASPLGLALAAERKALKESSEGGAQPQPPRPSPYEAP 1389
OY 356 TKTPKHSP-IKEEPC-----GSLSEYCKRE-----LRSGE-----TPEKRRSV--DTP 397
DB 1390 PTPHHSHANHEPVLRLMGASPPDPARRELGYRAGLSQSKSLPASPPAARRSLHMLP 1449
OY 398 PR-----LSTPQKGPSTHPK-----EKAFFSEIEDLPYLSTTEMYLCRWHQ----- 438
DB 1450 PTAQGVGPLLQLTGEPAPAPHPGVSKPWRSAAPPEPRLP-LHVRFLNCGPRAPVYSGR 1508
OY 439 -----PPSP--LPLRESSPKKEE 455
DB 1509 GPPSEDPGVPVPPSPRRSVPPSPSPRASE 1538
RESULT 10
ZAP3_MOUSE STANDARD: PRT; 1386 AA.
ID ZAP3_MOUSE


```

Db 146 GTAESGEAARAIPRE-----LPVSSSLPEIKPEHRCGLPNHFNGRAEGSRSELG 200
OY 159 --ASPATASD-----PAGPPL-----PL 177
Db 201 RAAGRGASDADGKLRNHFVGRSTYTKVLPAPKPHVNLNLTGKNDLSGHNHPF 260
OY 178 PCPPPLAPATAGTLAASEGRWKSRLKSPGGGGSGASQAACILQIILLDLLEQOQ 237
Db 261 SOPVHKGNTATKISLFENKRTKSSPRHNDIRGPRNTPASSKT-----FVGRAK 308
OY 238 QOLAKKEKEIEELK-----SERDILLARIEHMERMOVLKDD-----274
Db 309 LNLAKKAKEMQPEKKVMPNSPQNGVLVKTALIEKVVSEELIPATRGMDSENDA 368
OY 275 -----NEKER-----HKLFGYEETEEREETELSKILECOP 306
Db 369 LGPQNDQKADVQVDAGCLSEPVASALIPVKDHL-----EKEDSEADSKSLVE- 421
OY 307 ELSETSQTLP-----PKPFCGRSGKHKRSPGSTERKTPVKKLAPFSK 353
Db 422 NVTDAQDIPPTVDTKDLPTAMPKQHTFSDSQSPASSPCPSLSLSAP---AP--GD 475
OY 354 VKTKPKHSPTKEECGSL-----SETVCKRELRSGETPEKPSV--DTPPLST---PQ 404
Db 476 VPKDTCVOSPSSFPCTDLKVSNNKGCVLVPSRQNNKEMPLELGGETTPPLSTERSE 535
OY 405 KGPSTHPKE-----KAFSEIEDLPLST-----TEWLCRMHQPSPPLRLRESS 450
Db 536 AVSGECPRLVQVNSFLVPVESTODVSSQVTPESSEVREVLPCHSNEPVSVASCA 595
OY 451 PKKEETVA-----RCLMPSS-----VAGETSVLAVPSW 478
Db 596 PPOEEVLNENSHCTAELAAKSGQVIPPASEKTLPIQAQSGSRTPLMAESSPTNPS 655
OY 479 RDHSVEPLR-----DPNPSDLLE-----NLDVSFYSRHKLEDEKRRKMDQRIRE 527
Db 656 GNLHLPQRPDQVTNODSPASLTINISAGSDSVFSSS-----DKEKETE 702
OY 528 ORILQRLQ-----LRMKKKGIOESEPEVTSEFPEDVESLMTFLP-VVAFG-----R 577
Db 703 --IIQMDSAYCMPPKRRKAKRPNP--APHPAMPPIHEDHLEKVFDFKVFGLGKKKE 758
OY 578 PLPKLTPQNFELPWIDERSRCH 599
Db 759 SQEPMSPALHLMQNDTKSLR 780

RESULT 7
WASP_MOUSE STANDARD: PRT; 520 AA.
ID WASP_MOUSE
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein homolog (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c;
RX MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemhieu V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sf) mutation on
RT the X chromosome.";
RL Genomics 29:471-477(1995).
-1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION
CC (BY SIMILARITY).

```

```

CC -1- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U54788; AAC52556.1; -.
CC MGD; MGI:105059; Was.
CC InterPro: IPR000095; Pakbox/Rhoindng.
CC InterPro: IPR000697; Ranbpl_WASP.
CC InterPro: IPR001960; WH1.
CC InterPro: IPR003124; WH2.
CC Pfam; PF00568; WH1; 1.
CC Pfam; PF00786; PBD; 1.
CC Pfam; PF02205; WH2; 1.
CC SMART; SM00285; PBD; 1.
CC SMART; SM00461; WH1; 1.
CC SMART; SM00246; WH2; 1.
CC PROSITE; PS50108; CRIB; 1.
CC Repeat.
CC DOMAIN 41 147 WH1.
CC FT 240 253 CRIB.
CC REPEAT 354 363 GRSGLPPXP MOTIF 1.
CC FT 393 402 GRSGLPPXP MOTIF 2.
CC DOMAIN 162 167 POLY-PRO.
CC FT 314 321 POLY-PRO.
CC DOMAIN 324 341 POLY-GLY.
CC FT 368 373 POLY-PRO.
CC DOMAIN 376 379 POLY-PRO.
CC FT 384 390 POLY-PRO.
CC DOMAIN 397 403 POLY-PRO.
CC FT 408 424 POLY-PRO.
CC DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).
CC FT 520 520 ASP/GLU-RICH (ACIDIC).
CC SQ SEQUENCE 520 AA; 54191 MW; 9C223733C9FC68A CRC64;

Query Match
Best Local Similarity 26.1%; Pred. No. 0.04;
Matches 80; Conservative 29; Mismatches 106; Indels 91; Gaps 12;

OY 15 AGNPEQRUDYERAAALGPEDEPGAAR--HFLPRHKIKLEPGPPPLASSOGCGSPAPSPA 72
Db 276 AGISEAQLTDAETSKLITDFIEDOGLEAVRQEMRRQBPPLPPPCRCGGCGGGGGGG 335
OY 73 GCGGKRGRLLPAGAAPGOQEESSMGSVLP-----CPPATKQAGIGEPMAAGACS 126
Db 336 GGGGGGQPLRPVVS-----NKGSRGPLPPVPGAPPPPTPG-----PPPRGRGP 384
OY 127 PRPKYQAVLPLOTGSLVAANEPTPWADKG-----GAASPAATASDPAGPPPLPG 179
Db 385 PPP-----PPPATGRSGPPPLPGAGGPPAPPPPPPPPPPPCG 425
OY 160 -----PPPLAPRATAGTLAASEGRWKSRLKSPGGGGSGASQAACILQIILLDLLEQ 235
Db 426 SGPAPPPLPPPPVSGSPAP-----GGGRGA-----LLDQ 455
OY 236 OOOOQO-----AKKEIEELKSERDTL--ARIEMRRMOVLKQNEKERHKLFGYE 287
Db 456 IROGIQLNKTGCALENSVQGPAAQSGELVGNALMIVMKRSRVITHSDGED---QICE 511
OY 288 TEERE 293
Db 512 DEEDE 517

RESULT 8
WASP_MOUSE

```

```

FT VARSPLIC 532 582 VSANSSVITPRRMEDEPPKPKVSVPLRPITTSPP
FT VPOEPAPVPS -> PLSOSLLPMTLQSLSGOMAAPTTS
FT ACIDSPMLSPRLRPPCLTQLOL (IN ISOFORM
FT VARSPLIC 583 2715 MISSING (IN ISOFORM TRUNCATED).
FT CONFLICT 834 834 K -> E (IN REF. 6).
FT CONFLICT 941 941 S -> Y (IN REF. 6).
FT CONFLICT 1317 1317 E -> Q (IN REF. 6).
FT CONFLICT 1362 1362 H -> Y (IN REF. 6).
FT CONFLICT 1438 1438 D -> N (IN REF. 6).
FT CONFLICT 1918 1920 PLA -> GTR (IN REF. 5; AAH09337).
FT CONFLICT 2541 2543 DEE -> ARG (IN REF. 5; AAH07353).
FT CONFLICT 2622 2622 D -> H (IN REF. 6).
SQ SEQUENCE 2715 AA: 293511 MW: C0615B981BBE7BF CRC64;

Query Match 6.1%; Score 196.5; DB 1; Length 2715;
Best Local Similarity 21.0%; Pred. No. 0.13; Indels 285; Gaps 36;
Matches 158; Conservative 61; Mismatches 248;

QY 28 AALGPEDEPGAHAHFLPRHRLKPEGPPLASSOGSPAPSPAGCGKGR-----79
DB 3 AAGGGGCGPGSGSARGF-----PGRPRGAGGG-----GGGGRNGAERYVA 46
QY 80 -----GILLPAGAPGQ-----QESWGGSVP 101
DB 47 LRGGGATGPGGEGEDTALLRLRLRLRLRLMAGPRVQGRGRGRGWSGWS--104
QY 102 LPCEPATKQAGIGGEPAAAG-----AGCSPR-----P 129
DB 105 RGVPEEESDGEDEEFGFHSDEVAASSLSALRSORGRAPKRGKHKTTPLPP 164
QY 130 KYAVLPIQGTSLVAAAKEPTPMAGDK-----GAASPAATASDPAGP-172
DB 165 RLADVAP-----TPKTPARKRGEGETERMVQATLRLRAQAQAPRSRACEPTPR 217
QY 173 -----PP-----LPLP-GPPPLAFTATAGTLASEGRKMSMR-203
DB 218 RSRGRPPGRPAGPCRRKQAAVVAEAAVTIPKPEPPPPVP-----VKHOTGSMKEG 271
QY 204 -----KSLPGGGGSGAS-----SOAACTKQIILLDLT-IEQQQ 238
DB 272 PGCPGTPRRGGGSSRGGRGRGRGGLPVTKEVSR-----KVKMGQSLLESGQG 328
QY 239 QLAKEKEIEBELKSERDTLARI-----ERMRRLQVLKDKNEKHKLFQGYET 288
DB 329 QGQHEE-----SMQVPOKRVGSGGSCWKQKQKDLDEBEKKKEBEKKKEBEK 381
QY 289 EBEETELSEKIKLECOPELSETQTLPRKPFSCGSGKSGKHKRSFGSTERKTPV-KKL 347
DB 382 EERVAE-----EMMPAAKEKEAKLPPPLTPPASPPLPP--STSPPLPP 432
QY 348 APESKVKTKPKPKSPIKE-----PGSLSETVCKRELRSQETPEK-----PRSS 393
DB 433 PPPVSPPLPSPPPPPOEEDESPVPVATCSKRGKRPPLTISQRAERAPABGT 492
QY 394 VDPPLRLST-----POKGPSTHPEKKAASSEIED--LPYLSTTEMYLCR-----WHOPPP 441
DB 493 SPPTPTSTATGCPEDSPVAPAKSTELKNIQFIQVYVARSBSVLIKTPRRMDPDP 552
QY 442 SPLPLRESSPKKETVARCLMPSVAGETSVLAVPSMRDHSVEFLRDPNPSDLLENLDDS 501
DB 553 KP-PKEVSP-----VLKRPITTSPPVPOEPAPVPS-----PPRAPTPPS-----591
QY 502 VFSKRAKLELDEKRRKRWIDIORIREORLQRLQRLMYKKKKIQEESPEVTSFPEPDDV 561
DB 592 -----TPVPLPEKRRS-----ILREPTFKMTSL--TRLEPPPPPA--PPPPPA 630
QY 562 ESLMTPLPLVAVAGRP-----PKLTPQNFEL 589
DB 631 PS-----PPAPATSSRRPLLLRAPOFTSEAH 659

```

```

AIM1_HUMAN
ID AIM1_HUMAN STANDARD; PRT; 1723 AA.
AC 09Y4KL; 000296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Absent in melanoma 1 protein.
GN AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBT; TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Wistow G., Su Y.A., Meltzer P.S., Trent J.M.;
RT "AIM1, a novel non-lens member of the betagamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -1- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U83116; AAB53792.1; -.
DR EMBL; U83115; AAB53791.1; -.
DR HSSP; P02526; IGAM.
DR Genew; HGNC:356; AIM1.
DR MIM; 601797; -.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00030; crystall; 6.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR PRINTS; PR01367; BGCRYSTALLIN.
DR SMART; SM00458; RICIN; 1.
DR SMART; SM00247; XTALDg; 6.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 7.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
KW Repeat; Lectin.
FT DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.
FT DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.
FT DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.
FT DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.
FT DOMAIN 1405 1497 CRYSTALLIN BETA/GAMMA-LIKE 5.
FT DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.
FT DOMAIN 1586 1719 RICIN B-TYPE LECTIN.
SQ SEQUENCE 1723 AA: 188646 MW: 7550F681A627B09 CRC64;

Query Match 6.0%; Score 193.5; DB 1; Length 1723;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 164; Conservative 102; Mismatches 269; Indels 267; Gaps 38;

QY 6 AVEFAAAP-----AGNPDQRLQDERAAALGG-----PPDEPGAHAHFLPRHRLKPEP 55
DB 38 AVFDEVAAPNADSNASAEKVKVSPRALDGVASASAPSPKSPGKGL-RGESDRSK 96
QY 56 GPPLASSOGSPAPSPAGCGGKGRGL-LPAGAPGGOEESWGGSVP--LPCEPATKQA 112
DB 97 QP-----PASSPKRRGSRKALEAVAPAPASGRARA-KESPCKRVPDSSPYTK--145
QY 113 GIGGEPAAGAGCGSPREKYAVLPIQGTSLVAAAKE-----PTPWAG-DKGG-----158

```


RESULT 4
 ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
 AC 000268: Q99721: Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription Initiation factor TFIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4 OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=97336072; PubMed=9192867;
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;
 RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 mammalian cells.";
 RT Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggett C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levanthal M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McIlroy K., McMurtry A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swan R.M., Symcote N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942982;
 RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIID
 complex: hTAFII30 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y11354: CA72189.1; -;
 DR EMBL: AL137077: CAC36006.1; -;
 DR EMBL: AL109911: CAC22312.2; -;
 DR EMBL: U75308: AAC50901.1; -;
 DR TRANSFAC: T02328; -;
 DR Genew: HGNC:11537; TAF4.
 DR MIM: 601796; -;
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39
 FT DOMAIN 52
 FT DOMAIN 98
 FT DOMAIN 142
 FT DOMAIN 268
 FT DOMAIN 331
 FT DOMAIN 680
 FT DOMAIN 808
 FT DOMAIN 828
 FT CONFLICT 105
 FT CONFLICT 136
 FT CONFLICT 185
 FT CONFLICT 233
 FT CONFLICT 293
 SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CAC64;
 Query Match 6.1%; Score 197; DB 1; Length 1083;
 Best Local Similarity 20.8%; Pred. No. 0.051;
 Matches 155; Conservative 75; Mismatches 262; Indels 252; Gaps 30;
 QY 11 AAAPA-----GCNPGORDYERRAALGCPDEBPAAKHFPLPRHKKLEKPPPLASSGCGS 66
 DB 208 AAAPAVSLVNGPRAALLPLPKPAAPGTVIQTPEFVGAAAPADAPADSPPAAPAPAAA 267
 QY 67 PAPSAGCGGKGRGLLPAG---AAP-----CGQEESSGWSVPLPCPPPATQAGTIGE 117
 DB 268 PPPPPRPAALPAPRPGHPAPPPAPPAAPVPPPAANONGSAGAAPAPAP-PAAGGPAVCSQ 326
 QY 118 PAAAGACSPRPXYQAVLPITQGSVLAARKEPTPMAGDKGAASPAATAS----- 167
 DB 327 PGCAAAAPAPGKAPSPK---VVOAA-----PPAAQTAAAGSPASTASMTVIGPTMCG 379
 QY 168 ---DPAGPPPLPLPGPPPLAPATAGTGLASBGRKWSMRKSPGCGGSGSASSQAACLK- 223
 DB 380 ALDSPAAPVP-PAPGTPGLPKGAAAVT-----GSLSTPTATTSIGIRATLTPVYLAP 432
 QY 224 -----QILLQLDLIPQOQOQLOAKER----- 245
 DB 433 RLPOPPNPNTINQFPLPGMVLVRSNGGLMIPQALANQMAQNAHQPTMAPRAPAT 492
 QY 246 -----EIEELKSERDTLAP-----TERNERRNQLVKKNEKER----- 279
 DB 493 PPSAPVQISTVQAPGPTIARQVTPPTTIKQVSAQOTVQPSATLQSRPGVPOVLVIG 552
 QY 280 -----HKLFGYEREEHEEELSKILEQPELSEHSQTLPPKPF 351
 DB 553 AAOPTASIGTAVQGTGPTQTPGATTTSSAETMENVA-KCKNFSL---TL---IK 603
 QY 322 CGRSKGKHKRSPGSPGSEKTPYKLL-----APESKYKTKTPKSPKPEPCGS 371
 DB 604 LASSGK-----QSTETANVKELVONLDDKIEADPTSLVYELNSP---QP--- 649
 QY 372 LSETVCKREL-----RSQETPERKSSVDT----- 396
 DB 650 YLVPLFKRSLPALROLTPDSAAFLOQSQOQPPPTSQATATLAAVLSSSVQRTAGKTA 709
 QY 397 -----PPLSTPPO-----KGPST-----HKKKAFFSEIDIPLSTTEWYICR 435
 DB 710 TVNSALQPVLSLTPQVGVGKOGPTPLVIOQPPKPALIRP---PQVTLTQPPVMA 765

Db 917 -KEDAKPEKTEVAK--KEPDADA 938

RESULT 3

NFH_RAT

ID NFH_RAT STANDARD; PRT; 831 AA.

AC P1684; 063368;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Neurofilament triplet H protein (200 kDa neurofilament protein)

DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).

GN NFH OR NFH.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OC NCBI_TaxID=10116.

OX

RN

RP SEQUENCE FROM N.A.

RC TISSUE-Brain;

RX MEDLINE=89065087; PubMed=3143606;

RA Breen K.C., Robinson P.A., Wilson D., Anderson B.H.;

RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).

RL Identification of putative phosphorylation sites.";

RL FEBS Lett. 241:213-218(1988).

RN

RP SEQUENCE OF 37-831 FROM N.A.

RX MEDLINE=88309090; PubMed=2457365;

RA Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,

RA Jolles P.;

RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and

RT in situ detection.";

RL Blochem. Biophys. Res. Commun. 154:1099-1106(1988).

RL

RN

RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.

RX MEDLINE=87080760; PubMed=2878828;

RA Robinson P.A., Wilson D., Anderson B.H.;

RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide

RT (NF-H).";

RT FEBS Lett. 209:203-205(1986).

RL

RN

RP SEQUENCE OF 318-831 FROM N.A.

RX MEDLINE=89184647; PubMed=2928342;

RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,

RA Smolovitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;

RT "Cloning of a cDNA encoding the rat high molecular weight

RT neurofilament peptide (NF-H): developmental and tissue expression in

RT the rat, and mapping of its human homologue to chromosomes 1 and

RT 22.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).

RL

CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,

CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.

CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT

CC OBSERVED BY THE TWO SMALLER NF PROTEINS.

CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS

CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS

CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF

CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE

CC OF AXONAL CALIBER.

CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING

CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND

CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.

CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783

CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb.ch/announce/>)

or send an email to license@isb-sib.ch.

CC

DR EMBL: M37227; AAA41693.1; ALT_FRAME.

DR EMBL: X13804; CAA32038.1; ALT_FRAME.

DR EMBL: M21864; AAA41695.1; -.

DR EMBL: J04517; AAA41692.1; -.

DR PIR: A30796; A30796.

DR PIR: A25649; A25649.

DR PIR: B25649; B25649.

DR PIR: S02003; S02003.

DR InterPro: IPR001664; IF.

DR Pfam: PF00038; Filament; 1.

DR ProSite: PS00226; IF; 1.

DR Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.

FW NON_TER 1 1

FW DONAIN 1 1

FW CONFLICT 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.

FW CONFLICT 164 164 L -> I (IN REF. 2).

FW CONFLICT 185 185 I -> S (IN REF. 2).

FW CONFLICT 193 193 L -> T (IN REF. 2).

FW CONFLICT 199 199 M -> T (IN REF. 2).

FW CONFLICT 346 346 K -> N (IN REF. 1).

FW CONFLICT 373 373 A -> V (IN REF. 1 AND 4).

FW CONFLICT 482 482 G -> E (IN REF. 2 AND 4).

FW CONFLICT 485 485 P -> S (IN REF. 2).

FW CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).

FW CONFLICT 591 591 A -> V (IN REF. 4).

FW CONFLICT 727 727 A -> V (IN REF. 2 AND 4).

FW CONFLICT 757 759 AAP -> GST (IN REF. 4).

FW CONFLICT 769 769 T -> L (IN REF. 2).

FW CONFLICT 775 775 R -> P (IN REF. 2 AND 4).

FW SEQUENCE 831 AA; 89486 MW; 180973C3F13EF768 CRC64;

SO

Query Match 6.1%; Score 197.5; DB 1; Length 831;

Best Local Similarity 21.6%; Pred. No. 0.038;

Matches 116; Conservative 89; Mismatches 239; Indels 93; Gaps 22;

QY 11 A A A P A G C N P E O R L Y E R A A A L G P E D E G A E A H L P R H R K - - - L K E P G P L A S S O G S 66

DB 315 A K S P A - - - - - E K S P A E V K S P A E A K S P A E A K - S P A E V K S P A T V S G P A K S P A E V K S 365

QY 67 P A - - P S P A G C G K G R G L L P G A A P G Q G E E S W G S V P L P C P P P T K Q A G I G E - P A A A G A 123

DB 366 P A E V K S P A E A K S P A E A K S P A S V K S P G E A K S P A E A K S P A T V S P A T V S P A E A K S P A E V K S 425

QY 124 G C S P R K Q A V L P I O T G S L V A A K E P R W A G D G G A A S P A T A S D P A G P P L P G P P L 183

DB 426 P V T V K S P A E A K S P E V K S - P A S V K S P S E - A K S P G A K S P A E A K S - - - - - P V A K S P A 475

QY 184 A P T A G T L A A S E G R W K S M R K S P L G G G G S G A S S O A C L K Q I L L D L D L B Q Q Q O L O A K 243

DB 476 E A K S P A G A K P P A E A K S P A E A K S P A E A K S P A E A K S P V - - - E V K S P E A K S P V E G 532

QY 244 E K E I E L K S E N D T L A R L E R K R K Q L Y K N E K E R H L F G Y E T E R E E L S E K I K L E 303

DB 533 A K S L A E A K S P E A K A - - - - - K S P V E E I K P P A E V S P E A K S P M K R E A K - - 574

QY 304 C O P E L S E N S Q L P P - - K P F S G R S G K G H K R K S P G S T E R - T P Y K K L A P E R S K Y K T P K 360

DB 575 - S P E K A K T L D V K S P A P A R - - - P A K E A K R A D I R S P Q V A K S P A E A K S E K E T R E K 629

QY 409 T H - P R E K A F S S E I D L P L S T T E Y L C R W H O P P S P L P L R - E S S P K K E E T V A R C L M P S S V 466

DB 690 T E K P K D S G E A K E E - - - - A K E R K A A P E E T P A K I G V K E A K P K E A D A K A K E S K P 744

QY 467 A - - - - - G E T S L A V P S W R D H S V E P L R D N P D L L E N D S V F S R H A K L E I D E K R R K 518

DB 745 S E K E P K K E E V P A P E K D T K E E K T E S - - - - - K R E E K P M E A K A K E 787

FT CONFLICT 492 492 L -> S (IN REF. 1).
 FT COINCIDENT 670 670 I -> M (IN REF. 1).
 SQ SEQUENCE 1039 AA; 117412 MM; 4759E9B5E6E9F14 CRC64;

Query Match 6.58; Score 211; DB 1; Length 1039;
 Best Local Similarity 20.38; Pred. No. 0.012;

Matches 96; Conservative 91; Mismatches 171; Indels 114; Gaps 18;

QY 181 PPLAPYTAGTGLAASGRGSMKRSPLGGGSGGASOACLOIIL-----LQDLIE 234
 DB 569 PHLPRVAVAKVKTSTRESLETKA-----MTADKDAPOAKVIANHSTOTIDPVK 621
 QY 235 QQQOQLOAKKEI-----EELKSRDT-LIARIERMRRLVKKDKERHKL 282
 DB 622 TORLOVKIRQYEMPMRGTSSAPSDIRKQNVDPSTPTKTKSKILVNDKKTSET 681
 QY 283 FQGVETPERET---ELSEKIKLCOPELSETSTLP-----KPPSCGSGK 327
 DB 682 SOSPDOEIVETVRKRLAEHLKKEKLSQSHSOVTLKKIRERVATNLIYPP----- 733
 QY 328 GHRKSPFGSTERTKPYKLAPESSKVTTPKH---SPIKEPFGSLSETVCKRELRSQ 384
 DB 734 ---SAPVST---TTPATPTPTTPGSTPGHATVSSMDQELSAKSKAKAEQI--- 783
 QY 385 ETPEKPRSSVDTPRUSTPOK---GSTHPKEKAFSSIED---LPYLSIT--- 429
 DB 784 ATPLTPQSNSSVSSTSTIRKTLNCSPTHYSKATARSGLQSFRTATPTPTSTRTWEDQ 843
 QY 430 -----EMYLGRMHQPPSPPLPRESKPEETVARLMSVAGETSVLAIVPMRHSV 483
 DB 844 EPHCDNEFFL-----EEA-----DELLADNPSLETPKWRDPV 876
 QY 484 EPLRDPNPDLLENLDVSFKRAKLEDEKRRKMDIORIRQRILOLRMYKKG 543
 DB 877 PPSSDKIDTEL---LSQATFERRHQKVKVDEVRKCDAYKMEQILDLRMRANDV 933
 QY 544 IQSESP-EVTSFPEPDVDSLMTTPPLPVVAGRLPKLTPQ-NFELPMD 593
 DB 934 LVALDPLRASTFYPLPEDIAIOFVNEVTVQAGENVNMEARDDFGVMD 985

RESULT 2

NFH_HUMAN STANDARD; PRT; 1020 AA.

AC P12036;
 DT 01-OCT-1989 (Ref. 12, Created)
 DT 01-OCT-1989 (Ref. 12, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 DE (Neurofilament heavy polypeptide) (NF-H).
 GN NEFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE=88328981; Pubmed=3138108;
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Iazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 RT subunit (NF-H) and the gene encoding it.";
 RL EMBO J. 7:1947-1955(1988).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X15306; CAA33366.1; -
 DR EMBL: X15307; CAA33366.1; JOINED.
 DR EMBL: X15308; CAA33366.1; JOINED.
 DR EMBL: X15309; CAA33366.1; JOINED.
 DR PIR: S00979; QFNUH.
 DR Genew: HGNC:7737; NEFH.
 DR MIM: 162230; -
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT DOMAIN 1 100
 FT HEAD.
 FT DOMAIN 101 413
 FT ROD.
 FT DOMAIN 414 1020
 FT TAIL.
 FT DOMAIN 101 132
 FT COIL 1A.
 FT DOMAIN 133 145
 FT LINKER 1.
 FT DOMAIN 146 244
 FT COIL 1B.
 FT DOMAIN 245 266
 FT LINKER 12.
 FT DOMAIN 267 288
 FT COIL 2A.
 FT DOMAIN 289 292
 FT LINKER 2.
 FT DOMAIN 293 413
 FT COIL 2B.
 SQ SEQUENCE 1020 AA; 111780 MM; 117CGDCB3DCE1DA CRC64;

Query Match 6.18; Score 199; DB 1; Length 1020;
 Best Local Similarity 23.68; Pred. No. 0.04;
 Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

QY 90 GQOESWGSVPLPCPPATKQAGISGEPAAAGCSPPRYQAVPI-OTGSLVAAKE 148
 DB 489 GEEDEAGGEEETKSP-----AEBAAS-----PEKEKSPVKEAKSPAEKS 532
 QY 149 PTPWAGDKGGAASPAAT-----ASDPA-----GPPPLPLPGPPPLATATAGTGLAASGR 198
 DB 533 P-----EKEAKSPAEVKSPEKAKSPAEKSPAEKSPAEKSPAEKSPAEKSP 587
 QY 199 WKMKRSPPLGGGSGGASQAACLOILLQDLIEQOQOQLOAKKEIETELKSERDTLL 258
 DB 588 AKKEAKSPAEKSPAEKSPVK-----EEAKSPAEKSPVKEAKSP----- 629
 QY 259 ARIERMRMQLVKDKERHKLFOGYETEEETELSKIKLECOPELSETSTQTLPPK 318
 DB 630 AEVKSPEK-----AKSTPKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPV 684
 QY 319 PFCGSGSGKHKRSPGSTERTKTPYKLAPESSKVTTPKHSPKIEE-PCGSLSETVC 377
 DB 685 K---AEAKSPAEKSPV-KEAKSPAEKSPVKEAKSPAEKSPAEKSPAEKSPV 740
 QY 378 KRELRSQ---TEKPRSSVDTPRUSTPOKSGSTHP-----KKAKSSEIEDLPYSTTE 430
 DB 741 KEAKSPAEKSPAEKATTLVKSPEAKTPKEAKSPADKFPKAKSPVKEAKSPAEK 800
 QY 431 MYLCRMHQPPSPPLRE---SPKKEE---TVARCLMPSSVAGETSVLAIVPSW---RDHS 482
 DB 801 SPLKADAKAEKELPKKEEVKSKYKEEKQEVKVEPKKAEKAPATPKTEKKDKS 860
 QY 483 VE---PLRDPNPDLLENLDVSFKRAKLEDEKRRKMDIORI---REQRILOLRM 538
 DB 861 KEAPKPEAKPKPVKEEKPEAVPEKESKVE---AKKEAEDKKKVPTEKAPAKVEV-- 916
 QY 539 YKKKGIOSEPTVTSFPEPDVDS 563

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 14, 2003, 11:47:07 ; Search time 24 Seconds
(without alignments)
1061.103 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238
Sequence: 1 MTRSAVFKAAPAGNPE.....RSRCRLTIQKQTPHRTCRK 614

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211	6.5	1039	MSL1_DROME	P50535 drosophila
2	199	6.1	1020	NFH_HUMAN	P12036 homo sapien
3	197.5	6.1	831	NFH_RAT	P16884 rattus norv
4	197	6.1	1083	T2D3_HUMAN	O00268 homo sapien
5	196.5	6.1	2715	TRX2_HUMAN	O9um66 homo sapien
6	193.5	6.0	1723	ATM1_HUMAN	O9y4k1 homo sapien
7	192.5	5.9	520	WASP_MOUSE	P73315 mus musculu
8	190	5.9	1087	NFH_MOUSE	P13246 mus musculu
9	182.5	5.6	2161	SHK1_HUMAN	O9y566 homo sapien
10	181.5	5.6	1386	ZAP3_MOUSE	O9y017 mus musculu
11	178.5	5.5	2167	SHK1_RAT	O9y448 rattus norv
12	174	5.4	755	RRE1_HUMAN	O92766 homo sapien
13	173	5.3	351	CSR_PLAKU	P04922 plasmodium
14	172.5	5.3	707	SFRP0_HUMAN	P23246 homo sapien
15	172	5.3	3969	HRX_HUMAN	O03164 homo sapien
16	171	5.3	777	RNN1_RAT	O64548 rattus norv
17	171	5.3	1647	SN24_HUMAN	P46634 homo sapien
18	169.5	5.2	616	VGE_HUMAN	O15240 homo sapien
19	169.5	5.2	857	NFM_CHICK	P16053 gallus gall
20	169	5.2	268	CDX1_MOUSE	P16111 mus musculu
21	168.5	5.2	1183	DRPL_RAT	P34258 rattus norv
22	168.5	5.2	3530	MY15_HUMAN	O9ukn7 homo sapien
23	167	5.2	2142	BAT2_HUMAN	P46634 homo sapien
24	166	5.1	2220	CAB1_HUMAN	O9y6f0 homo sapien
25	165	5.1	358	CERB4_RAT	P05534 rattus norv
26	165	5.1	1362	BRD4_HUMAN	O60885 homo sapien
27	165	5.1	1509	GSR1_HUMAN	O9uzm4 homo sapien
28	164.5	5.1	757	TAU_HUMAN	P10636 homo sapien
29	164	5.1	806	MR07_MOUSE	O9yvs8 mus musculu
30	164	5.1	1618	NRST_HUMAN	P46681 homo sapien
31	163	5.0	502	WASP_HUMAN	P42768 homo sapien
32	163	5.0	1217	AR4_MOUSE	O88573 mus musculu
33	163	5.0	2472	NCR2_MOUSE	O9wu42 mus musculu

34	162.5	5.0	1300	SAL3_HUMAN	O9bxa9 homo sapien
35	162.5	5.0	1446	IE18_PRYKA	P33479 pseudorale
36	162	5.0	817	VRP1_YEAST	P37370 saccharomyc
37	161.5	5.0	758	HUNB_DROME	P05084 drosophila
38	161.5	5.0	1262	CA13_CHICK	P12105 gallus gall
39	161.5	5.0	1461	IE18_PRYF	P11675 pseudorale
40	161	5.0	579	SOC7_MOUSE	O8vhaq2 mus musculu
41	161	5.0	1394	CNG4_BOVIN	O28181 bos taurus
42	161	5.0	2182	CAB1_RAT	O88480 rattus norv
43	160.5	5.0	1185	DRPL_HUMAN	P54259 homo sapien
44	160	4.9	1130	AB11_HUMAN	P00519 homo sapien
45	159.5	4.9	359	CERB4_MOUSE	P35366 mus musculu

ALIGNMENTS

```

RESULT 1
MSL1_DROME STANDARD: PRT; 1039 AA.
AC P50535;
DT 01-OCT-1996 (Rel. 34, created)
DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 01-OCT-1996 (Rel. 34, last annotation update)
DE Male-specific lethal-1 protein.
GN MSL-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE OF 85-1039 FROM N.A.
RC STRAIN=Canlon-S;
RX MEDLINE=93314941, PubMed=8325488;
RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the X chromosome in
RT males";
RT Genetics 134:545-557(1993).
[2]
RN REVISIONS, SEQUENCE FROM N.A.
RP MEDLINE=95300219, PubMed=7781064;
RA Kelley R.L., Solovayeva I., Lyman L.M., Richman R., Solovay V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
RT regulators on the X chromosomes and female lethality in Drosophila.";
RT Cell 81:867-877(1995).
-!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
-!- TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
-!- DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MSL-1 AND
-!- MSL-2 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
-!- PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
-!- BINDING.
-!- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
-!- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
-!- OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
-!- AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
-!- SITES IN MALES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL: L42514; AAA98918.1; -
FlyBase: FBgn0005617; msl-1.
Nuclear protein.
CONFLICT 188 193 PUPPA -> HCHLIP (IN REF. 1).

```



```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129 SVJ:
RX MEDLINE-98345363; PubMed-9679147;
RA Dieck S., Sammarti-Vila L., Langeneese K., Richter K., Kindler S.,
RA Soyke A., Mex H., Smalla K.H., Kampf U., Franzer J.T., Stumm M.,
RA Garner C.C., Gundelfinger E.D.;
RT "Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively
RT localized at the active zone of presynaptic nerve terminals.";
RL J. Cell Biol. 142:499-509(1998).
DR EMBL; Y17034; CAA76598.1; JOINED.
DR EMBL; Y17035; CAA76598.1; JOINED.
DR EMBL; Y17036; CAA76598.1; JOINED.
DR EMBL; Y17037; CAA76598.1; JOINED.
DR EMBL; Y17038; CAA76598.1; JOINED.
DR MGD; MGI:1277955; Bsn.
SQ SEQUENCE 3942 AA; 418739 MW; 1502678636C4DCB CRC64;

Query Match 6.1%; Score 197.5; DB 11; Length 3942;
Best Local Similarity 22.5%; Pred. No. 0.0033;
Matches 138; Conservative 62; Mismatches 173; Indels 241; Gaps 32;

QY 32 GGPEDPGAAEAHFLPRHRKLEKPPPLASSOG-----GSPAP----- 69
DB 2125 GGPD-----LVQYQPOH-----GPGLSAPQGLAPLRSGLGNPTYPEGQPSPCNLAY 2172
QY 70 SPAGCGKGRGLLP-----GAAPQ-----OES----- 95
DB 2173 GPASQATAVKQLLPSTAVRAADGMITSTINTPIATLPITTOPASVLRPMVKGMYRP 2232
QY 96 --WGG--SVPL-----PCP-----PRATKQAGIGEPD- 119
DB 2223 YVSGGVAVPLSLTRVPMIAPRVPLPAGLYRYPARPRPLASSVPAEGPYL-GRPA 2291
QY 120 --AAGAGCSPRKQAVPIPIOTGSLVAAAKEPTPWAGDKGCAASPAATASDPAGPPPL 177
DB 2292 TKASGAGPPEPPELVAGV-----AREEP-----STTAPAVIKKAPVAPAPGPA 2335
QY 178 PGPPPLAVTAVAGTLAASEGRMKSMKSPGLGGGGGASQAACLKQILLQLDLIEQQ 237
DB 2336 PAPP-----GQKPAGEA-----AAGSGSVLSRPAEKE-----EASQ 2369
QY 238 QOLAKKEIELEKSEDTLLARIERMERMQLVKKDKERHKLFGYETEEREETELS 297
DB 2370 EDROKQOE-QLQLERE---RYLEKLRQLRLQELERVEL-----QRHREEDLL 2419
QY 298 EKIKLECPPELSETSQTLPPKPFSCGRSGKH-----KKSPPGSTERKTPVKKLAE 350
DB 2420 -----VORELOEL-QTI-----KQHVLOQOOEEROQAFALOREQLAQRLOLE 2461
QY 351 FSKVTKTPKHSPIKEEPCGSLSEIVCKRELRSOETPEKPRSSVDTPPRLSTPOKGPSTH 410
DB 2462 -----QIQO-----LQQQLQLQLEEQKOKAKAPPATCEAPSRGPPPA 2499
QY 411 PEKAFSSEIEDLPYLSTTEMYLCRWHPSPPLRLRESSPKKEETVARCLMPSSVAGET 470
DB 2500 ATELAQNGQY--WPLTHAFLAVAGTGGPGP-----REPVLRGLPSS-ASDM 2546
QY 471 SVLAVPSN-----RDHSVEPLRD-----PNPSDLLENLDVSFSKRRAKLELDEK 515
DB 2547 SLQTEEQWEAGRSQIKRHSMPRLADACEPSGPDPSVTRIRIADSSVQTD-----DEE 2599
QY 516 RRRKMDIQRIRQR 529
DB 2600 GEGRYLVTRRRRTR 2613
```

Search completed: July 14, 2003, 11:50:38
Job time : 91 secs

DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE High molecular-weight neurofilament.
 GN NF-H.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91038277; PubMed=2230956;
 RA Chin S.S., Liem R.K.K.;
 RA "Transfected rat high-molecular-weight neurofilament (NF-H)
 RT coassembles with vimentin in a predominantly nonphosphorylated form.";
 RL J. Neurosci. 10:3714-3726(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Chin S.S.M., Liem R.K.H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 DR EMBL: AF031879; AAB87068.1; -
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 DR Coiled coil; Intermediate filament.
 SQ SEQUENCE 1072 AA; 115349 MW; 89A146D457AD78B CRC64;

Query Match 6.1%; Score 197.5; DB 11; Length 1072;
 Best Local Similarity 22.1%; Pred. No. 0.00082;
 Matches 112; Conservative 82; Mismatches 235; Indels 77; Gaps 20;

QY 11 AAAPAGGNPQRLDYERAAALGPEDEPGAEEHFLPRHKK-----LKEPPPLASSGGG 66
 DB 548 AEAKSPAEAKSPAEVKSPTATVKSPEAKSPAEAK-SPAELVKSPTATVKSPEAKSPAEAKS 606
 QY 67 PAISPAAGCGGKRGKLLPAGA-----APGQDEESVGSVLPCCPPATKQAGTIGE-P 118
 DB 607 PAEVKSPVEAKS-----PAEAKSPASVKSPEAKSPAEAKSPAEVKSPTATVKSPEAKSP 661
 QY 119 AAAGAGCSPRKQAVLPVPIOTGSLVAAXKEPTPMAGDKGAASPAATAASDPAGPPLPLP 178
 DB 662 AEVKSPTVKSPEAKSPVEVKS-PASVKSPE-ASPAAGAKSPAEAKSPVAKSPEAK 719
 QY 179 GPPPLAATAGTLAASEGKMSKSPKLGSGGSGASQAACIKQLLLDLIEQOOQ 238
 DB 720 SPAEAKPPAEKSPAEK-----KSPAFAKSPAEKSPAEKSPV---EVKSPAEK 768
 QY 239 QLOAKKEIEELKSERDTLLARIERMERRMOLVKKDKNEKRHKLFQGYETEEREETLSE 298
 DB 769 PVKEGAKSLAEKSPKKA-----KSPVKEIEIKPPAEVKSPEAKSPKKE 812
 QY 299 KIKLECPLESESTQLPRPKFSGSGRGKHKRSPSGSTER-KTPVKTLAPESKSKYTK 357
 DB 813 EAK----SPEAKTLIDVKSPEAKTPAK--EBAKRPADIRSPQYKSPAEKSPKEKSPKEETR 867
 QY 358 TPKHSPIKEBPCKSLSETVCK---RELRSQETPEKPRSSV-----DTPRLSTP---QK 405
 DB 868 TEKVAPEKKEVKSPEVEVKAKPEPKKVEEKEKTPATEVKEKSKXDAPAEAKPKAEK 927
 QY 406 GPSTH-PKEKAFSSIEDLPYLSTTEKYLCRMHOPPPSPPLR-ESSPKKEEYVARCLMP 463
 DB 928 EPLTEPKKDSPEAKKEE-----AKEKKAAPAEETPAKLGVEEAKPKAEKADAKAEK 982
 QY 464 SSVV-----GETSVLAVPSWRDHSVE 484
 DB 983 SKSESEKPKKEEVPAPAEKKDKTEE 1008

RESULT 14
 0921R1
 ID 0921R1 PRELIMINARY; PRT; 2157 AA.
 AC 0921R1;

DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE BAT2.
 GN NFRB11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
 RA Shaffer T., Ratcliffe A., Loretz C., Lasky S., Hood L.;
 RT "Sequence of the mouse major histocompatibility class III region."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF109719; AAC82480.1; -
 DR MGD; MGI:1340031; Nfrb11.
 SQ SEQUENCE 2157 AA; 229072 MW; 35B21F6B7C911F40 CRC64;

Query Match 6.1%; Score 197.5; DB 11; Length 2157;
 Best Local Similarity 23.5%; Pred. No. 0.0017;
 Matches 124; Conservative 49; Mismatches 213; Indels 141; Gaps 25;

QY 17 GNPEQRLDYERAA-----ALGG--PEDEPGAEEHFLPRHKKLEP-----GPPPL-- 59
 DB 496 GADPRLKAAPAAPVTPAPAPALPVYVKEIPAPAPALPPTPEKEPEEPAQAPPVQA 555
 QY 60 ASSGGGSPAP---SPAGCGGKRGK-----LLPAAAGGQDEESVGSVLPCCP 106
 DB 556 APSPGAVPVLTVSGGGCTANSSSGSFEASPYEPOLIPSGEGPEPE-----VPPPTTP 610
 QY 107 PATKQA-----GIGG--EPAAAGAGCSPRKQAVLP-----IOTGSLVAAXKEPTWAGD 155
 DB 611 PAPKMEKKGVGSTRTQPPSGIG--YKTKQSLPRPFRQOQOGLKQOQOQOQOQ 667
 QY 156 KCGAASPA-----ATASDPAGPPLP---LPG-----PPLAATATGTLAASE 196
 DB 668 QOQAPAPAPVPPSPPOVTLGAAPAPAPAPPPKALPGALGRPPMPV-----MNF 720
 QY 197 GRMKSM-----RKSPLG-----GGGSGASQAACIKQLLLDLIEQOOQLOA 242
 DB 721 PRMMMTIPPYDPLRLOGRPPLDTPPGVHSG-----LVPRESDSGSSSEP 768
 QY 243 KEKEIEELKSERDTLLARIERMERRMOLVKKDKNEKRHKLFQGYETEEREETE-LSEKIK 301
 DB 769 FERRAPPLRERGT-----PPVDPKLAW-----GDVFTTTPDPRPLTSLRQAD 815
 QY 302 LECOPLESETSQLPRPKFSGSGRGKHKRSPSGSTERKTPVKKLAP----- 349
 DB 816 EEEKSMRSETPPVPPPPPYLA--NYPGFENGTPGPPIISFPLEESAIPGPRPLPWP 873
 QY 350 -EFSKYTKTPKSPKPIKEEPCGSLSETVCKRELRSQETPEKPRSSVDTPEKPS 408
 DB 874 DEAAKQAPPPKPEKSEEPQOLSGPEAGKPARGGGPPPRRENTERTRW-----GPR 928
 QY 409 THPKKAFSSIEDLPYLSTTEKYLCRMHOPPPSPPLRRESSPKKE 455
 DB 929 PGSCRGRIPEEPGVPRRAGPI-----KKPPPVVEELPPKSLSE 969

RESULT 15
 088737
 ID 088737 PRELIMINARY; PRT; 3942 AA.
 AC 088737;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE Bassoon.
 GN BSN OR BASSOON.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;


```

Db      809 VSEPTSTANLGLLSRSPVCPRLPGPENHRAGSNEDPILASGTPPTIPDET 868
QY      127 ----PAPKY-----QAVLPIDG-----LVAAKEPPWADKGGAASPAATASD 168
Db      869 GGVNPPAVHYHKEASDVEISLESDSDSVYVEGLPPLPPPPSGATPPPIAPTCP 928
QY      169 PGPPLPL-----PGPPLAPATAGTLAASGRKWSMKSPGLGGGGS 213
Db      929 PTPSPVPAPKEEPELPAAPGPLPPPPPPVPVPGVTLPPQ-----LVPEGTGCG 981
QY      214 GASQAACLKQILLDLIDLEQQOQLOAKKEIEELKSEDTLLARIEMERRMQLVK 273
Db      982 GPALBEDLTVININSDEEEEGEGBEEEBEEDDEEE-----EEEDDFEEEBE 1036
QY      274 DNE---KERHKLFGYETE-EREETELSEKIKLECOPELSETSOTLPPKFGSGRGK 329
Db      1037 DEEEYEEEEEEEEEDEEEEGELEEEEEEDEEBEELEEV---EDLEGTAG--- 1090
QY      330 KRKSPGSTERKTPVKKLAPFESKVTYKTPKNSPIKEEPOGSLSETVCKRELKSOETPEK 389
Db      1091 -----GEVEGAPPPPTLPALPPESPCKVQPEPEPGLLEV-----EEGTE 1136
QY      390 PRSSVTPRLS-----TPQKGPSTHPKEKASSEIDLPYLSTTEMYLC 434
Db      1137 EERGAOTAPTLAEALPSOGEEVEREGESPAGP--PPOELVEEESXPPTLLEETEDG 1193
QY      435 RWHQPPSPPLPLRESSPKKEETVARCLMPSSVAGETSVLAPSMRHSVEPLRDPMSD 493
Db      1194 SDKVQPPPTPAEEF--METETAEALOEKEODDTAAMLA--DFIDCPPDEKPPPTTE 1248

```

RESULT 10

09UJ57

PRELIMINARY: PRT; 1026 AA.

```

ID 09UJ57 09UJ57 PRELIMINARY: PRT; 1026 AA.
AC 09UJ57 09UJ57 13, Created
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Neurofilament protein (K1A0845 protein).
GN K1A0845
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhu Y., Han Y.
RT "Molecular Cloning of human hSTE cDNA."
RL Pei-Ching I Ko Ta Hsueh Hsueh Pao 31:531-531(1999).
RN [2]
RP SEQUENCE OF 94-1026 FROM N.A.
RC TISSUE-BRAIN:
RX MEDLINE=9156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka K., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XII. the complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 5:355-364(1998).
CC - SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL; AF203032; AAF13722.1;
DR EMBL; AB020652; BAA74868.1;
DR InterPro; IPR001386; Histone_H1/H5.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PRO0624; HISTONEH5.
DR PROSITE; PS00226; IF; 1.
KM Coiled coil; Intermediate filament.
SQ SEQUENCE 1026 AA; 112479 MW; 0879B6A08D208C17 CRC64;

```

Query Match 6.28; Score 200; DB 4; Length 1026;
Best Local Similarity 22.48; Pred. No. 0.00055;

Matches 116; Conservative 69; Mismatches 228; Indels 104; Gaps 21;

```

QY      90 GOOEESMGSVPLPCPPATKQAGIGGPAAGAGCSPPKYOAYPI-OTGSLVAAKE 148
Db      489 GEEEAEGGEETKSP-----AEERAS-----PEKAKSPVKEAKSPREAKS 532
QY      149 PTPWADKGGAASPAAT-----ASDPA-----GPPPLPLPGPPLAPATAGTLAASGR 198
Db      533 P-----EKEEAKSPAVKSPREKAKSPAKDEAKSPPEKAKSPAEVAKSPREKAKSP 587
QY      199 WKSMSKSPGLGGGSSAQAACLKQILLDLIDLEQQOQLOAKKEIEELKSEDTLL 258
Db      588 AKKEAKSPAEKAKSPREKAKSPV-----EAKSPAEKAKSPVKEAKSP----- 629
QY      259 ARIEMERRMQLVKD-NKERHKLFGYETEEREETELSEKIKLECOPELSETSOTLPP 317
Db      630 AEYKSPKAKSPKEAKSPREKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSP 689
QY      318 KPESGSGKGRKSPFSTERKTPVKKLAPFESKVTYKTPKNSPIKEE-POGSLSETV 376
Db      690 VK---AEAKSPKAKSPV-KEAKSPKAKSPVKEAKSPKAKSPKAKSPKAKSP 745
QY      377 CKRELKSOE---TPKPRSSVTPPLSTPQKGPSTHP---KEKAFSEIDLPYLSTT 429
Db      746 VKKEAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSP 799
QY      430 EMYLCRWHQPPSPPLPLRESS-----PKKEETVARCLMPSSVAGETSVLAPSMRHS 482
Db      800 -----KSPERAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSP 851
QY      483 VEPLRDPNPSDLLENLDVSFSRRHAKLEDEKRRKRWMDIQRLEQRILQLRLMYKKK 542
Db      852 KAP-ATPKTEKKKSKKEAPKKEAPKPYEEKKEPA--VKKEKSVKAKKEADKDKK 908
QY      543 GIOSEPEVTSF-----PEPDVVS 563
Db      909 -VPTPEKAPAKVEKEDAKPKRETEVAKKEPDADA 944

```

RESULT 11

09R0R5

PRELIMINARY: PRT; 1034 AA.

```

ID 09R0R5 09R0R5 PRELIMINARY: PRT; 1034 AA.
AC 09R0R5 09R0R5 13, Created
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Transcription factor CA150b.
GN TCERG1 OR TAF2S.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimada M., Horiyome T.;
RC TISSUE-LIVER:
RX MEDLINE=20047056; PubMed=10578054;
RA Shimada M., Saito M., Katakai T., Shimizu A., Ichimura T., Omata S.,
RA Horiyome T.;
RT "Molecular cloning and splicing isoforms of mouse p144, a homologue of
RT CA150."
RL J. Biochem. 126:1033-1042(1999).
DR EMBL; AB023485; BAA86392.1;
DR MGD; MGI:1926421; Tcerg1.
DR InterPro; IPR002713; FF.
DR InterPro; IPR002965; P-rich_extensions.
DR InterPro; IPR001202; WW_Rsp5_WWP.
RP Pfam; PF01846; FF; 5.
DR Pfam; PF00397; WW; 3.

```



```

DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Unknown (Protein for IMAGE:4156419) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010457; AAL10457.1; -
FT NON_TER 1
SQ SEQUENCE 1052 AA; 112161 MW; C71A751164C8C1AB CRC64;

Query Match
Best Local Similarity 23.1%; Score 209.5; DB 4; Length 1052;
Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;

QY 55 PCPPPLASSOGG-----SPAPSPAGCGGKGRGLLPAGAAPG--QOEESSWGS 99
Db 622 PCPPPTANHLGLSVPLVSVPRLLPGRENHAGSNEDPLASGTPPTIPDETFTGGR 681
QY 100 VPLPCPPPAATKQ-----AGIGEPAAAGAGCSPRPKYQAVLP1Q 138
Db 682 VPRPAVNHVDKEASDVEISLESDDSVYIYVEGLPPLPPPPSGATPP----- 732
QY 139 TGSVLAANKPTPWADKGAAPATASD-----PAGPPPLPPLPPPLAATATGTLA 193
Db 733 ----LAPGPPPT-----ASPPVPAKEEPBELPAAPGLP-PPPPPPPPPGVTLTP 778
QY 194 ASEGKMKSKKSPGCGGSGASSQAACLKQILLDLLEQOOOLQAKEREIEELKSE 253
Db 779 PPG-----LVPEGTPGGGPPALBEDLTVININSSD--EEEEEEEEEEEEEEEE 829
QY 254 RDTLLARIERMERRMOLVKKNKRNKLFQGYETEEREETELSEKIKLECOPELSETSQ 313
Db 830 ED-----FEEEEEDEEYEFEEEEEFEFEFEFEFEFEFEFEFEFEFEFEFEFE 882
QY 314 TLPPKPSGSGKSGKHKRSPPFSTERKTPVKTLADEFKVKTKTKHSIRKEPGSLS 373
Db 883 ----EDLEFETAG-----GEVEGAPPPPLPPLPPSPKVPQEPPEPPGCLL 929
QY 374 E-----TWCKRELRSQ-ETPEKPRSSVDPTRPRLSTPQKPSYHPEKAF 416
Db 930 EWEPEGTBEERGADTAPTLAPALPQSGEVEREGESPAAGPPQELVEEPSAP--TLL 987
QY 417 SSEIEDLPYLSTEMTLCRMHOPPPSPPLIRESSPKKETVANCLMPSSVAGETSVLAVP 476
Db 988 EEEETED-----GSDKV-----QPPPE-TPAEEF--METETEALQEKEDDTAAMLA-- 1032
QY 477 SWRDHSEVPLRDPNPSD 493
Db 1033 DFTDCPPDDEKPPPPTE 1049

RESULT 6
O8TDF0 PRELIMINARY; PRT; 2429 AA.
AC O8TDF0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid inducible in neuroblastoma cells RAIND1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21902045; PubMed=11904404;

```

```

RA Merrill R.A., Plum L.A., Kaiser M.E., Clagett-Dame M.;
RT "A mammalian homolog of unc-53 is regulated by all-trans retinoic acid
RT in neuroblastoma cells and embryos."
RL Proc. Natl. Acad. Sci. U.S.A. 99:3422-3427(2002).
DR EMBL; AF466143; AAL96479.1; -
SQ SEQUENCE 2432 AA; 261380 MW; DAC37362819DE2DF CRC64;

Query Match
Best Local Similarity 22.7%; Score 203.5; DB 4; Length 2429;
Matches 133; Conservative 79; Mismatches 254; Indels 121; Gaps 23;

QY 6 AVFKAAPAA--GNPEQRLDYERAAALGPEDEFGAALHFLPHRRLKEPCPLASS 62
Db 209 AVSQVAGASQCAQAPQOQV-----PVTPOAPCQHPAPAHQOSKAO---AEM 254
QY 63 QGGSPPA---SPAGCGGKGRGLLPAGAAPGQOEESSWGSVPL--PCPPPTAKQAGIG 116
Db 255 QSRLEGPPTARVSAAGSEARTRGSGTANNRRSSFNNDKSKVTSPPPPSSHEK---- 310
QY 117 EPAAAGAGCSPRPKYQAVLP1QTSVLAANKPTPWADKGAAPG-----SPATASDPAG 171
Db 311 EPLASSASSHPGMSDNAPASLESSTFTNCSTSSAIPQGAATKPMRSKLSVYKHSAT 370
QY 172 PPLPLPGLPPLAATATGTLAASBGRKMSMK-----SPUGGGGSGASSQAACLKQI 225
Db 371 VSMLSVYKPPGPEAPRPTPEAMKPAANNOKSMLEKLLFNKSGSGKAGEPGSRDTSCERL 430
QY 226 LILQDLILGQOOOLQAKEREIEEL-----KSEBDTLARIEMERRMOLVKRD---NKR 277
Db 431 ETLP---SESEBELTASMLTTPVGPASSPPIAKGIAQKTFSSALTNKSKSLKGNK 487
QY 278 ERHKLQGYETEREE-----TELSEKIKLECOPELSETSQTPPKPFGSGRSG---KGH 329
Db 488 ERKQQRKQREKDKSKDLARASVTERLDLKEEKEDPSGAAYVEPMKSKSIASFIPKG 547
QY 330 K-----KSPGSGTERKTPVKKTLAPFSKYKTKPK--HSPKEPCGSLETVCKRLRQ 384
Db 548 KNSAKKEPAPSHSGIP---KPGKSMGKSPAPAPSK-----GERSHSGLSLG 597
QY 385 ETEKPE-----RSSVDPTRPRLSTPQKPSYHPEKAFSS-----IEDLPYLSTEMTL 433
Db 598 LPQAKQLDGRHSSSSSSSLASSGKPGGTTLNHSISQTVSSGVTTQTGSGNTVSYQL 657
QY 434 CRMHOPPPSPPLIRESSPKKETVANCLMPSSVAGETSVLAVSWDHSEVPL----- 486
Db 658 -----POPOQOQNHNPATAVAPFLVRSQDTEGANTAAESSGVSVEPSHFTKGTQ 708
QY 487 -----NDP-----NPSDLLENDDSVFSKR-----HAKLE 511
Db 709 PALEELTGEDPEARRLTVKNIADLRQNLLETWSSLRGTQVTHSTLE 755

RESULT 7
O8TDF0 PRELIMINARY; PRT; 2432 AA.
AC O8TDF0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid inducible in neuroblastoma cells RAIND1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21902045; PubMed=11904404;
RA Merrill R.A., Plum L.A., Kaiser M.E., Clagett-Dame M.;
RT "A mammalian homolog of unc-53 is regulated by all-trans retinoic acid
RT in neuroblastoma cells and embryos."
RL Proc. Natl. Acad. Sci. U.S.A. 99:3422-3427(2002).
DR EMBL; AF466144; AAL96480.1; -
SQ SEQUENCE 2432 AA; 261722 MW; 2666957ED0A726F9 CRC64;

```

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkios G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfanchko C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris M.P., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclab J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003660; AAF53689.1;
 DR Flybase: FBgn005617; msl-1.
 FT NON_TER
 SQ SEQUENCE 1039 AA; 117503 MW; 644C21C0390A2750 CRC64;

Query Match 6.6%; Score 213; DB 5; Length 1039;
 Best Local Similarity 20.9%; Pred. No. 9.5e-05;
 Matches 99; Conservative 85; Mismatches 172; Indels 118; Gaps 18;

181 PPLATATAGTTLAEGRRKMRKSPGLGGGSGASSQAACIKQILL-----LQDLIE 234
 DB PKHLPKAVAPKVTSTSSRESTLPKA-----NTADIDAPAQKIAHNSKOTDTDYK 621
 OY 235 QQQQOOLQAKKE-----IEELKSERDTLLARIEMRMQLVKKDKNEKERH 280
 DB 622 TORLQVKIRQYEMHDMRTGSSAPSDIRKQKNVDVSTPEKTIKSKMLV--NDKKTTS 679
 OY 281 KLFQCYETEEERET---ELSEKIKLECPLESEISQITP-----PKPPSCGRS 325
 DB 680 ETSOSPDEIDVEYTRRKLAEHLKELLQSQSHSSQVTLKIRERVATNLIYPP----- 733
 OY 326 GKGHRKRPFGSTERKTPVKKLAPEFSKVTKTPRH-----SPIKEPPGSLSEYVCKRRLR 382
 DB 734 -----SAPVST---TTPAPTPSTPTPGSTPOHANTSSMDOTISAKKSKRAEIO- 783
 OY 383 SQEPPEKRRSSVDPRLSTPOK-----GPSTHPEKAFSEIED-----LPVLTST-- 429
 DB 784 --APPLTPQSSSVSSTSTIRKTLNNSCPHTYSKATARSGLQSRFRATPEFYSTRWE 841
 OY 430 -----EMYLCKRHOPPPPLPLRESPPKEETVACLPSSVAGETSVLAIVSWSDH 481
 DB 842 DOEFHCNDEFL-----EEA-----DELLANPSLELPKMDV 874
 OY 482 SVEPLRDPNPSDDLLENLDVSFVKRAKLEDEKRRKMDIORIQRLOLRMYK 541
 DB 875 PVPSSSKDITDEL---LSDATFERRHQKYVDEVDKRCRDARIKMEQIRLEQLRKRRMD 931
 OY 542 KGIOSEBP-EVTSFEPDPDVESIMTTPFLPVAVAGRLPKLTPQ--NFLPMLD 593

DB 932 EVLVALDPLRASTFPLPDEDIEALQFVNEVTVQAFGENGVNMKEARDGCVWVD 985
 RESULT 4
 AC 09BU60 PRELIMINARY; PRT; 746 AA.
 ID 09BU60;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypoetical 79.2 KDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC002875; AA02875.1; -
 DR InterPro: IPR002965; P-Fich_extensions.
 DR PRINTS: PR01217; PRICHEXTENSN.
 KM Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 746 AA; 79187 MW; 35A186C5A7ADEA8B CRC64;

Query Match 6.5%; Score 209.5; DB 4; Length 746;
 Best Local Similarity 23.1%; Pred. No. 0.00011;
 Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;

OY 55 PGPPLASSQGG-----SPASSPAGCGGKRGILLPAGAARG--QOESWGS 99
 DB 316 PGPPTTANHLGLSVGLVSVPRLLPGFENRAGSNEDPILAPSGTPPTTPPDTEGGR 375
 OY 100 VPLPCPPRATKO-----AGIGEPAAAGAGCSPPRYOAVLPQ 138
 DB 376 VPRPAFVHYDKESADVEISLESDDSVIIVBGLPLPLPPPPSGAVPP----- 426
 OY 139 TGSLYAAAKETTPMAGDKGAASPAATASD-----PAGPPPLPLGPPPLATAGTLA 193
 DB 427 ---IAPGPPT-----ASPPVPAKKEPEELPAAPGLP--PPPPPPVPPQVPLP 472
 OY 194 ASEGKMRKMRKSPGLGGGSGASSQAACIKQILLQDLILQOQOOLQAKKEIEKSE 253
 DB 473 PPQ-----LVPESTPPGGCPPLALEDLIVININSSD--EEEEESEEEEEEEEEE 523
 OY 254 RDTLLARIEMRMQLVKKDKNEKRNKLFQYETEEERETELSEKIKLECPLESETSQ 313
 DB 524 ED-----PEEBEDEEEYFEEEBEEDFEFEEEBEEDBELEEBEEDDEEBEELLEV-- 576
 OY 314 TLPPKPFSCGSGCKHKKSPGSTERKTPVKKLAPEFSKVTKTPRH-----PKPPSCGRS 373
 DB 577 ---EDLBFGTAG-----GVEEECAPPPTPLPALPPESPAPKQVPPPEPEPGLL 623
 OY 374 E-----TWCKRELRSQ-EPEKRPSSVDTPRLSTPOKSGTHPKKAKF 416
 DB 624 EYEEGTEEGAGDAPLAPALPDSQGEVEREGSGPAGPPPOELVEEFSAPP--TLL 681
 OY 417 SSEIEDLPYLSTTEMYLCRMHOPPPSPPLRESPPKKEETVACLPSSVAGETSVLAIV 476
 DB 682 EEEETD-----GSDKV-----QPPPE--TPAEBE--METETAAALQEKEDODTAAMLA-- 726
 OY 477 SMRHSVPELPDNPDS 493
 DB 727 DFIDCPPDDEKPPPEPE 743
 RESULT 5
 ID 096FT1 PRELIMINARY; PRT; 1052 AA.
 AC 096FT1;

SQ SEQUENCE 616 AA; 67332 MW; 221C84C306C7D07D CRC64;
 Query Match 96.0%; Score 3110; DB 11; Length 616;
 Best Local Similarity 95.9%; Pred. No. 5e-177;
 Matches 591; Conservative 10; Mismatches 13; Indels 2; Gaps 2;

QY 1 MTRSAVFKAAAPAGNPNBQRLDYERAAALGGPEDEPGAENHFLPRHKKLEPPPLA 60
 1 MTRSAVFKAAAPAGNPNBQRLDYERAAALGGPEDEPGAENHFLPRHKKLEPPPLA 60
 DB 1 MTRSAVFKAAAPAGNPNBQRLDYERAAALGGPEDEPGAENHFLPRHKKLEPPPLA 60
 QY 61 SSGGSPAPSPAC-GGKGGGLLPAGAAAGQEESSGVPLPCPPPAKOGIGEPA 119
 61 SSGGSPAPSPAC-GGKGGGLLPAGAAAGQEESSGVPLPCPPPAKOGIGEPA 119
 DB 61 SSGGSPAPSPACGGGKGGGLLPAGAAAGQEESSGVPLPCPPPAKOGIGEPA 120
 QY 120 AAGAGSPRKYQAVLPVIGTSTVAAAKERTPMAGKGAAPAAATSPAPAPPLPLP 178
 120 AAGAGSPRKYQAVLPVIGTSTVAAAKERTPMAGKGAAPAAATSPAPAPPLPLP 178
 DB 121 AAGAGSPRKYQAVLPVIGTSTVAAAKERTPMAGKGAAPAAATSPAPAPPLPLP 180
 QY 179 GPPPLAPTATAGTLAASGGRKMSKPSLGGGGSGASSQAACLKQILLQLDLIEQOOQ 238
 179 GPPPLAPTATAGTLAASGGRKMSKPSLGGGGSGASSQAACLKQILLQLDLIEQOOQ 238
 DB 181 GPPPLAPTATAGTLAASGGRKMSKPSLGGGGSGASSQAACLKQILLQLDLIEQOOQ 240
 QY 239 QLOAKEKEIEELSERDTLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSE 298
 239 QLOAKEKEIEELSERDTLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSE 298
 DB 241 QLOAKEKEIEELSERDTLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSE 300
 QY 299 KIRLECOPELSETSOQLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKT 358
 299 KIRLECOPELSETSOQLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKT 358
 DB 301 KIRLECOPELSETSOQLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKT 360
 QY 359 PKHSPIKEEPCGSLSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSS 418
 359 PKHSPIKEEPCGSLSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSS 418
 DB 361 PKHSPIKEEPCGSLSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSS 420
 QY 419 EIEDLPLSTETMYLCHWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSW 478
 419 EIEDLPLSTETMYLCHWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSW 478
 DB 421 EIEDLPLSTETMYLCHWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSW 480
 QY 479 RDSHVEPLRPNPSDLLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLM 538
 479 RDSHVEPLRPNPSDLLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLM 538
 DB 481 RDSHVEPLRPNPSDLLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLM 540
 QY 539 YKKKGIOSEPEVTSPFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRC 598
 539 YKKKGIOSEPEVTSPFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRC 598
 DB 541 YKKKGIOSEPEVTSPFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRC 600
 QY 599 RLEIOKKOTPHRTCRK 614
 599 RLEIOKKOTPHRTCRK 614
 DB 601 RLEIOKKOTPHRTCRK 616

RESULT 2
 Q9DSC9 PRELIMINARY; PRT; 370 AA.
 ID Q9DSC9
 AC Q9DSC9
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 4930463F05R1K protein.
 GN 4930463F05R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7B/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Iwawa K., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK015496; BAB2968.1; -
 DR MGI:1922153; 4930463F05R1K.
 SQ SEQUENCE 370 AA; 43184 MW; 2347FA88F39B5654 CRC64;

Query Match 52.8%; Score 1711; DB 11; Length 370;
 Best Local Similarity 89.3%; Pred. No. 3.4e-94;
 Matches 324; Conservative 10; Mismatches 13; Indels 16; Gaps 1;

QY 252 SERDTLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSEKIRLECOPELSET 311
 252 SERDTLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSEKIRLECOPELSET 311
 DB 24 TNRKKLLARIERMERMOVLKKNKKERHKLFGYETEEEREPELSEKIRLECOPELSET 83
 QY 312 SOTLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKPKHSPIKEEPCGS 371
 312 SOTLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKPKHSPIKEEPCGS 371
 DB 84 SOTLPRKPSGSGSGKHKRSPGSTERKTPVKKLAPESKVKTKPKHSPIKEEPCGS 143
 QY 372 LSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSSIEDLPLSTTEM 431
 372 LSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSSIEDLPLSTTEM 431
 DB 432 LSETVCKRELRSOETPEKPNSSVDTPRLSTPOKGPSTHPEKAFSSIEDLPLSTTEM 444
 QY 432 YLCRMHQPSPPLRLRESSPKKEETV-----ALPSRDSHVEPLRPNP 491
 432 YLCRMHQPSPPLRLRESSPKKEETV-----ALPSRDSHVEPLRPNP 491
 DB 204 YLCRMHQPSPPLRLRESSPKKEETV-----ALPSRDSHVEPLRPNP 247
 QY 492 SDLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLMYKKKGIOSEPEV 551
 492 SDLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLMYKKKGIOSEPEV 551
 DB 248 SDLENLDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLMYKKKGIOSEPEV 307
 QY 552 TSFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRCLEIOKKOTPHRT 611
 552 TSFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRCLEIOKKOTPHRT 611
 DB 308 TSFPEPDVDSLITPFLPVVAFGRPLKLTPONFELPWLDERSRCLEIOKKOTPHRT 367
 QY 612 CRK 614
 612 CRK 614
 DB 368 CRK 370

RESULT 3
 Q9VJ66 PRELIMINARY; PRT; 1039 AA.
 ID Q9VJ66
 AC Q9VJ66
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 GN MSL-1 OR CG10385.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:10 ; Search time 87 Seconds
(without alignments)
1454.172 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3338

Sequence: 1 MTRSAVFKAAPAGNPPE.....RSRCRLKQKQPHRCRK 614

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_protent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvivirus.*
- 16: sp_bacteriophage.*
- 17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	96.0	616	11	Q9CXF9 mus musculus
2	1711	52.8	370	11	Q9DSC9 mus musculus
3	213	6.6	1039	5	Q9VU56 drosophila
4	209.5	6.5	1052	4	Q9BU60 homo sapien
5	209.5	6.5	1052	4	Q96FT1 drosophila
6	203.5	6.3	2429	4	Q8TDF0 homo sapien
7	203.5	6.3	2432	4	Q8TDF0 homo sapien
8	200.5	6.2	1021	4	Q8TDF0 homo sapien
9	200.5	6.2	1021	4	Q8TDF0 homo sapien
10	200	6.2	1026	4	Q9U571 homo sapien
11	200	6.2	1034	11	Q9U571 homo sapien
12	200	6.2	1034	11	Q9U571 homo sapien
13	197.5	6.1	1072	11	Q9U571 homo sapien
14	197.5	6.1	1072	11	Q9U571 homo sapien
15	197.5	6.1	1072	11	Q9U571 homo sapien
16	196.5	6.1	821	4	Q96JW3 homo sapien

17	194.5	6.0	1123	11	Q9DBD5 mus musculus
18	193.5	6.0	520	11	Q61078 mus musculus
19	193.5	6.0	1173	11	Q9QXV3 mus musculus
20	193.5	6.0	1197	11	Q9QXV3 mus musculus
21	193	6.0	753	4	Q9H987 homo sapien
22	192.5	5.9	762	4	Q9C0H9 homo sapien
23	192.5	5.9	1475	5	Q9N4G4 caenorhabditis
24	192.5	5.9	5085	11	Q9JRK5 mus musculus
25	192	5.9	3534	12	Q9Z266 equine
26	192	5.9	3536	5	Q9VZ30 drosophila
27	191	5.9	540	4	Q96B87 homo sapien
28	190.5	5.9	917	4	Q9P1Y5 homo sapien
29	190	5.9	1060	3	Q9C259 neurospora
30	190	5.9	1098	4	Q14776 homo sapien
31	189.5	5.9	599	6	Q9SK67 macaca fasc
32	189.5	5.9	606	6	Q28687 oryctolagus
33	189	5.8	1154	4	Q9P2P0 homo sapien
34	189	5.8	1431	4	Q8WXP9 homo sapien
35	189	5.8	5147	4	Q9Y6V0 homo sapien
36	189	5.8	554	4	Q9HBO2 homo sapien
37	188.5	5.8	675	11	Q91Y49 mus musculus
38	188.5	5.8	1800	4	Q9H1C1 homo sapien
39	188	5.8	1862	4	Q9H1C2 homo sapien
40	188	5.8	2703	5	Q9VEG7 drosophila
41	188	5.8	2715	5	Q9VEG7 drosophila
42	188	5.8	966	4	Q8WU64 drosophila
43	187.5	5.8	1644	4	Q96JX5 homo sapien
44	187.5	5.8	1644	4	Q96JX5 homo sapien
45	187.5	5.8	2701	4	Q9Y520 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9CXF9	PRELIMINARY:	PRT:	616 AA.
AC	Q9CXF9	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	4121402D02R1K protein.			
GN	4121402D02R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batlov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bafsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker G., Winding L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK014463; BAB29369.1;			
DR	MGD; MGI:1921276; 4121402D02R1K.			

Db	123663	ATTCAGTTTCGAGTGGAATGCCCGCCGACAAATTCGCCAAAACAAGTGTG-----TTTCCTT	123716
OY	4147	AACCAAGTTTAAGCTAGCTGTTAATAAATAAAA---AAATTGCTTGTCTGTCACTTCA	4203
Db	123717	AATAAAGTTTCACCTAAGTGTGTGAATAATCCAAAACAAATGCTAATCTGTCCACTTCA	123776
OY	4204	GCATTGTTTAATGCCCACTTTCATATATGTTGTCTGTGTGAATTAATACCTTTGTATAC	4263
Db	123777	ACTTTCGTTTAATGCCCACTTTCATATATGTTGTCTGTGTGAATTAATACCTTTGTACACCA	123836
OY	4264	ATTTCGTGATGCTAAATATGTTGCTTGTGAATAATCTTAATAAGATTCAATTTGTAAT	4323
Db	123837	ATTTCGTATGAAGAAATGTTCAATGTCCTGTG-AGATTCCTACGAAGATTCATATGTAGGT	123895
OY	4324	AAACATATGCTGGCTTAATAAAAAATAAAAAATAAAAA 4358	
Db	123896	AAATTAATGTGGCTGTCTAATTTTGTGAATAAACAAAA 123930	

```
Search completed: July 14, 2003, 16:44:42
Job.time : 10978 secs
```


Template: 10 ng
Primer: each 5 µM
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/µl
Total Vol: 20 µl

Buffer:
MgCl₂: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

FEATURES
source
Derived from dbEST (genbank accession R78466).

Location/Qualifiers
1..240
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="319.5 CR from top of Chr17 linkage group"

STS
primer_bind 1..150
primer_bind 1..23
complement(127..150)
BASE COUNT 76 a 45 c 51 g 68 t
ORIGIN

Query Match 5.5%; Score 240; DB 11; Length 240;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2943 AAAATATCAGATAGACATCCACGACCTTGTCTCACTGTAAGATCCCTTTACTA 3002
Db 1 AAAATATCAGATAGACATCCACGACCTTGTCTCACTGTAAGATCCCTTTACTA 60
OY 3003 TCTGAAGAGAAATAGACCCACGACCTTGTCTCAATATATAGAAATGCTTTCTT 3062
Db 61 TCTGAAGAGAAATAGACCCACGACCTTGTCTCAATATATAGAAATGCTTTCTT 120
OY 3063 AGCTTCAGACATATGTGTGAAACAAAGTAGGGGTCTAATCTCTTGAAGTAGGGGCT 3122
Db 121 AGCTTCAGACATATGTGTGAAACAAAGTAGGGGTCTAATCTCTTGAAGTAGGGGCT 180
OY 3123 TTATCTTAAAGAGATATGTCCTCCAGATTATTAGCACTTTAGAGAGAGCAAGGT 3182
Db 181 TTATCTTAAAGAGATATGTCCTCCAGATTATTAGCACTTTAGAGAGAGCAAGGT 240

RESULT 14
AL354659 135060 bp DNA linear PRI 15-NOV-2001
LOCUS Human DNA sequence from clone RP11-157G15 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL354659 AC024007
VERSION AL354659.14 GI:16972840
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 135060)
AUTHORS Griffiths, C.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humgen@esanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:1390411.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-157G15 is from the library RP01-11.1 constructed by the group of Pletier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-157G15. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-157G15 is at 135060 in this sequence. The true left end of clone RP11-291H23 is at 34561 in this sequence. The true right end of clone RP11-361K17 is at 2000 in this sequence.

FEATURES
source
Location/Qualifiers
1..135060
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-157G15"
/clone_1b="RP01-11.1"

BASE COUNT 44915 a 24717 c 23719 g 41709 t
ORIGIN

Query Match 5.4%; Score 235.4; DB 9; Length 135060;
Best Local Similarity 81.8%; Pred. No. 4.2e-34;
Matches 323; Conservative 0; Mismatches 61; Indels 11; Gaps 4;

OY 3968 TTTTGTGAGAGAGAGTGGAGAGTTCAGTATAGCAAGTCCAAAGTCAAAATTTACAGT 4027
Db 126046 TTTGTGAGAGAGAGAGATGGAAGTATAGTAATGTTCCAAAGTGAAGTACAGT 126105
OY 4028 GTTAAAGTGTGGGGGAAATTAATGTTATTTTCCCTACATGGGATACAACTGTG-A 4086
Db 126106 GTTAAAGTGTGGGGGAAATTAATGTTATTTTCCCTACATGGGATACAACTGTG-A 126165
OY 4087 ATTCAATCTCACTGAAGGCGCTGCAGATTCCTAAACATATGTTGTTTCTT 4146
Db 126166 ATTCAATCTCACTGAAGGCGCTGCAGATTCCTAAACATATGTTGTTTCTT 126219
OY 4147 AACCAAGTTTAAAGCTAGTGTAAATTAATTAATAA--AAATGCTTGTCTGTACTTCA 4203
Db 126220 AATTAAGTTTAAAGCTAGTGTAAATTAATTAATAA--AAATGCTTGTCTGTACTTCA 126279
OY 4204 GCTTTGTTTAAAGCTATTTCAATTTGTTGTCTGTGTGTAATCAATTTTGATAC 4263
Db 126280 AATTTGTTTAAAGCTATTTCAATTTGTTGTCTGTGTGTAATCAATTTTGATAC 126339
OY 4264 ATTTGATGTGTAAGAAATGTTGTTGTAATTAATTAATTAAGAGTCAATTTGAT 4323
Db 126340 ATTTGATGTGTAAGAAATGTTGTTGTAATTAATTAATTAAGAGTCAATTTGAT 126398
OY 4324 AAATTAATTTGCTGTAAATAAATAAATAAATAA 4358
Db 126399 AAATTAATTTGCTGTAAATAAATAAATAAATAA 126433

RESULT 15
AL513172 142908 bp DNA linear HMG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-291H23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 17 unordered pieces.
ACCESSION AL513172

*	16888	19027:	contly of 2140 bp	in length
*	19028	19127:	gap of 100 bp	
*	19128	22490:	contly of 4363 bp	in length
*	23491	23590:	gap of 100 bp	
*	23591	25951:	contly of 2361 bp	in length
*	25952	26051:	gap of 100 bp	
*	26052	29248:	contly of 3197 bp	in length
*	29249	29348:	gap of 100 bp	
*	29349	32751:	contly of 3403 bp	in length
*	32752	32851:	gap of 100 bp	
*	32852	37466:	contly of 4613 bp	in length
*	37465	37564:	gap of 100 bp	
*	37565	42592:	contly of 5028 bp	in length
*	42593	42692:	gap of 100 bp	
*	42693	49347:	contly of 6655 bp	in length
*	49348	49447:	gap of 100 bp	
*	49448	54633:	contly of 5186 bp	in length
*	54634	54733:	gap of 100 bp	
*	54734	65288:	contly of 7555 bp	in length
*	65289	62388:	gap of 100 bp	
*	62389	146836:	contly of 84450 bp	in length
*	146839	146938:	gap of 100 bp	
*	146939	158635:	contly of 1697 bp	in length
*	158636	158735:	gap of 100 bp	
*	158736	175149:	contly of 16408 bp	in length
*	175144	175243:	gap of 100 bp	
*	175244	200086:	contly of 24843 bp	in length
*	200087	200186:	gap of 100 bp	
*	200187	207214:	contly of 7028 bp	in length

```
misc_feature      1. .9177
                  /note="assembly_fragment"
clone_end:Sp6
vector_side:left"
misc_feature      9278. .9931
                  /note="assembly_fragment"
misc_feature      10032. .10809
                  /note="assembly_fragment"
misc_feature      10910. .13046
                  /note="assembly_fragment"
misc_feature      13147. .14912
                  /note="assembly_fragment"
misc_feature      15013. .16787
                  /note="assembly_fragment"
misc_feature      16888. .19027
                  /note="assembly_fragment"
misc_feature      19128. .23490
                  /note="assembly_fragment"
misc_feature      23591. .25951
                  /note="assembly_fragment"
misc_feature      26052. .29248
                  /note="assembly_fragment"
misc_feature      29349. .32751
                  /note="assembly_fragment"
misc_feature      32852. .37464
                  /note="assembly_fragment"
misc_feature      37565. .42592
                  /note="assembly_fragment"
misc_feature      42693. .49347
                  /note="assembly_fragment"
misc_feature      49448. .54633
                  /note="assembly_fragment"
misc_feature      54734. .62288
                  /note="assembly_fragment"
misc_feature      62389. .146838
                  /note="assembly_fragment"
misc_feature      146935. .158635
                  /note="assembly_fragment"
```

Query Match	5.5%;	Score 242;	DB 2;	Length 207214;
Best Local Similarity	89.7%;	Pred. No. 2.4e-35;		
Matches 260;	Conservative 0;	Mismatches 30;	Indels 0;	Gaps 0;

[illegible]

RESULT	13			
G22752				
LOCUS	G22752	240 bp	DNA	linear
DEFINITION	human STS WI-17575, sequence tagged site.			STS 31-MAY-1996
ACCESSION	G22752			
VERSION	G22752.1	GI:1343078		
KEYWORDS	STS; STS sequence; primer; sequence tagged site.			
SOURCE	Homo sapiens STS derived from sequences in dbEST and the Unigene collection.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 240)			
TITLE	Hudson, T.			
JOURNAL	Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS			
COMMENT	Unpublished (1995)			

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AAAATATCAGATAGACATCCCA
Primer B: ACTGTGTTTCACACATATGCTG
STS size: 150
PCR Profile:

 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
PCR Cycles: 35
Thermal Cycler:

Protocol:

QY	3914	AGGACCTCTATTATTATTTGCCCCCTCTTAGAGTTAATCTCCCTTGAATCTTGTT	3973
Db	429	AGGACCTAATATTGATTATCTCCTCTCTAG-----CTGCTTNTGGTTTGGAGATGCT	375
QY	3974	GAGAGGAGGTTGGACAGTAGATTAGCAAAAGTTCCAAAGTGCMAAATTAACAGTGTGAGA,	4033
Db	374	GAGGTTGTTGAGGAGGAGGTTGGAGCAAAAGTTCCAGAGAGCGAAGCAGCATGTGT---	318
QY	4034	GTTGGGGGGGAAAATTAGTCTTATTTTTCCTACATGGGATACAACTGTG---AATTC	4090
Db	317	GAGGTGTGGGGAATTATTTATTTTCCCTACATGGGATACAACTGTGTAGATTC	258
QY	4091	AATCTTAACTGA-----AGGCCCTGCAGTCTCCTAAACATAGTT-GTGTGTTTTCT	4144
Db	257	AGCCTTCAACAGAAAGAGAGGCCCTGCAGTCTCCTAAACATAGTTCCTACTTTCT	198
QY	4145	TTAACAAGATTAAAGTAGTGTATATAATTA--AAAATGCTGTCTGTCTACTTC	4202
Db	197	TTAACAAGTTAAAGTAGTGTATATAATTAATTAAGAAAGAAAGTCTGTCTCCACCCTC	138
QY	4203	AGCTTGTATTATGCCATTTCAATATGTTGTCTGTGTGAATTCATTAATTTGATAC	4262
Db	137	AGCTTGTATTATGCCATTTCAATATGTTGTCTGTGTGAATTCAT-ACATTTGATAC	79
QY	4263	CATTTCGATGTGAATAATGTTGTCTGTGAATATCTTAAAGAGTTCAATTTGAAA	4322
Db	78	CATTTCGATGTGAATAAGTCGCTGTGTGAATATCTTAAAGAGTTCAATTTGAAA	19
QY	4323	TAACTATTGSGCTGTT	4340
Db	18	TAACTATTGSGCGTT	1

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted 224-JAN-2002 Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 207214)
Birden, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boughalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazarro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S., Fardo, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gargyala, S., Ginde, S., Gord, S., Guyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Larocque, K., Lamaizares, R., Landers, T., Lehoczeky, J., Levine, J., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menais, L., Mihov, T., Mleenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Reiter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schpack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Trivis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2002 this sequence version replaced gi.18308342.
All repeats were identified using RepeatMasker:
Smith, A.F., A. & Green, P. (1996-1997)

RESULT	12
AC107643	
LOCUS	
DEFINITION	AC107643 Mus musculus clone RP23-54L14, WORKING DRAFT SEQUENCE, 21 ordered pieces.
ACCESSION	AC107643
VERSION	AC107643.2 GI:20258435
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Mus musculus.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 207214)
AUTHORS	Mus musculus, clone RP23-54L14
TITLE	Unpublished
JOURNAL	2 (bases 1 to 207214)
REFERENCE	Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
AUTHORS	Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

TITLE

Direct Submission

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genomic Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 207214)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhagbater,B., Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeBariello,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,N., Gage,D., Galagan,J., Gargyna,S., Ginde,S., Gord,S., Goyette,M., Gisham,L., Grand-Pierre,N., Hagens,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Larocque,K., Lamazares,R., Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Milhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thumann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiljev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,N.T., Young,G., Zainoun,J., Zemlek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genomic Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2002 this sequence version replaced gi:18308342.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L14249
Center clone name: 54_L14
----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 20014 bases at least Q40
Consensus quality: 203403 bases at least Q30
Consensus quality: 204607 bases at least Q20
Insert size: 210000; agarose-1p
Insert size: 205214; sum-of-contigs
Quality coverage: 7.0 in Q20 bases; agarose-1p
Quality coverage: 7.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submittor.

This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 9177: contig of 9177 bp in length
* 9178 9277: gap of 100 bp
* 9278 9931: contig of 654 bp in length
* 9932 10031: gap of 100 bp
* 10032 10809: contig of 778 bp in length
* 10810 10909: gap of 100 bp
* 10910 13046: contig of 2137 bp in length
* 13047 13146: gap of 100 bp
* 13147 14912: contig of 1766 bp in length
* 14913 15012: gap of 100 bp
* 15013 16787: contig of 1775 bp in length
* 16788 16887: gap of 100 bp

24653 24752: gap of unknown length
24753 26187: contig of 1435 bp in length
26188 26287: gap of unknown length
26288 27567: contig of 1280 bp in length
27568 27567: gap of unknown length
27668 28980: contig of 1313 bp in length
28981 29080: gap of unknown length
29081 30136: contig of 1056 bp in length
30137 30236: gap of unknown length
30237 31632: contig of 1396 bp in length
31633 31732: gap of unknown length
31733 33270: contig of 1538 bp in length
33271 33370: gap of unknown length
33371 35465: contig of 2095 bp in length
35466 35565: gap of unknown length
35566 37201: contig of 1636 bp in length
37202 37301: gap of unknown length
37302 38947: contig of 1646 bp in length
38948 39047: gap of unknown length
39048 40710: contig of 1663 bp in length
40711 40810: gap of unknown length
40811 42463: contig of 1653 bp in length
42464 42563: gap of unknown length
42564 43912: contig of 1349 bp in length
43913 44012: gap of unknown length
44013 45591: contig of 1579 bp in length
45592 45691: gap of unknown length
45692 47620: contig of 1929 bp in length
47621 47720: gap of unknown length
47721 48829: contig of 2109 bp in length
48829 49929: gap of unknown length
49930 51465: contig of 1536 bp in length
51466 51565: gap of unknown length
51566 52865: contig of 1300 bp in length
52866 52965: gap of unknown length
52966 54343: contig of 1378 bp in length
54344 54443: gap of unknown length
54444 55895: contig of 1452 bp in length
55896 55995: gap of unknown length
55996 57574: contig of 1579 bp in length
57575 57674: gap of unknown length
57675 59301: contig of 1627 bp in length
59302 59401: gap of unknown length
59402 61393: contig of 1992 bp in length
61394 61493: gap of unknown length
61494 63075: contig of 1582 bp in length
63076 63175: gap of unknown length
63176 64817: contig of 1642 bp in length
64818 64917: gap of unknown length
64918 67402: contig of 2485 bp in length
67403 67502: gap of unknown length
67503 69777: contig of 2275 bp in length
69778 71982: gap of unknown length
71983 72082: contig of 2105 bp in length
72083 74289: contig of 2207 bp in length
74290 74389: gap of unknown length
74390 76937: contig of 2548 bp in length
76938 77037: gap of unknown length
77038 79003: contig of 1966 bp in length
79004 79103: gap of unknown length
79104 81519: contig of 2416 bp in length
81520 81619: gap of unknown length
81620 84864: contig of 3245 bp in length
84865 84964: gap of unknown length
84966 87539: contig of 2575 bp in length
87540 87639: gap of unknown length
87640 89437: contig of 1798 bp in length
89438 89537: gap of unknown length
89538 92108: contig of 2571 bp in length
92109 92208: gap of unknown length
92209 94902: contig of 2694 bp in length
94903 95002: gap of unknown length

* 95003 98082: contig of 3080 bp in length
Query Match 8.5%; Score 371; DB 2; Length 193189;
Best Local Similarity 96.2%; Pred. No. 6.8e-60;
Matches 380; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 1059 GGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGTCTGTTAAAGCTG 1118
DB 185340 GTACTTAATAGAAAGAAAGCCATTGGAATACAGAAAGAAAGTCTGTTAAAGCTG 185281
OY 1119 GCTCTGAATTTTCAAAAGTCAAAACAAATCTCTAGACACTCTCTATTAAAGAGAA 1178
DB 185280 GCTCTGAATTTTCAAAAGTCAAAACAAATCTCTAGACACTCTCTATTAAAGAGAG 185221
OY 1179 CCTGTGTTCTTATCTGAAGTCTGTTGAAGTGAATGAGAGCAAGAAACCCA 1238
DB 185220 CCTGTGTTCTTATCTGAAGTCTGTTGAAGTGAATGAGAGCAAGAAACCCA 185161
OY 1239 GAAAGCCCGGTCTTCAAGTGAACACCCCAAGACTCTCCACTCCCAAGGAGACC 1298
DB 185160 GAAAGCCCGGTCTTCAAGTGAACACCCCAAGACTCTCCACTCCCAAGGAGACC 185101
OY 1299 AGCACCCTCCCAAGAGAAAGCCCTCTCAAGTGAATGAAGATTGCTTACTTCC 1358
DB 185100 AGCACCCTCCCAAGAGAAAGCCCTCTCAAGTGAATGAAGATTGCTTACTTCC 185041
OY 1359 ACCACAGAAATGATTTGTGTTGGCAGCAGCCCTCCCATCAGCTTACATTAAGG 1418
DB 185040 ACCACAGAAATGATTTGTGTTGGCAGCAGCCCTCCCATCAGCTTACATTAAGG 184981
OY 1419 GAATCCTCTCCAAAGAGAGAGACTGTACAG 1453
DB 184980 GAATCCTCTCCAAAGAGAGAGACTGTACAG 184946
RESULT 11
AX400340 590 bp DNA linear PAT 06-JUN-2002
LOCUS Sequence 16 from Patent WO2010453.
DEFINITION AX400340
ACCESSION AX400340
VERSION AX400340.1 GI:21336520
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
Elashoff,M.R.
TITLE Molecular toxicology modeling
JOURNAL Patent: WO 0210453-A 16 07-FEB-2002;
Gene Logic, Inc. (US)
FEATURES
source location/Qualifiers
1..590
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/note="EMBL/GenBank Accession No. AA799599"
BASE COUNT 205 a 134 c 106 g 143 t 2 others
ORIGIN
Query Match 6.3%; Score 274.8; DB 6; Length 590;
Best Local Similarity 79.2%; Pred. No. 6.1e-42;
Matches 442; Conservative 0; Mismatches 94; Indels 22; Gaps 9;
OY 3794 GTTGTGCTTACCCAGGAGGAGTCCCATGTTGACTTGAAGTAGACTGAAAGATCC 3853
DB 547 GTTCTGCTTATCAGAGAGAGGAGTCTCATATCTG-TTTNGCTAGACTGAGAGAAATCC 489
OY 3854 ACGAGGCTTATCAGAGAGAGGAGTGAAGTGAATGCTGTTGCTGCTGCTGCTGCTG 3913
DB 488 ATGAGGCTTATCAGAGAGAGGAGTGAAGTGAATGCTGTTGCTGCTGCTGCTGCTG 430

OY 1225 GCCAAGAAACCCAGAAAGCCCGGTCTTCAGTGGACACCCACCAAGACTCTCCACTC 1284
 |||||||
 Db 511 GCCAAGAAACCCAGAAAGCCCGGTCTTCAGTGGACACCCACCAAGACTCTCCACTC 570
 |||||||
 OY 1285 CCCAAAGGAGCCAGCAGCCCATCCCAAGAGAAAGCTTCTCAAGTAGATAGAACTT 1344
 |||||||
 Db 571 CCCAAAGGAGCCAGCAGCCCATCCCAAGAGAAAGCTTCTCAAGTAGATAGAACTT 630
 |||||||
 OY 1345 TCCCGTACCTTTCACACAGAAATGTATTGTGTCGTGGACAGCCCTCCCATAC 1404
 |||||||
 Db 631 TCCCGTACCTTTCACACAGAAATGTATTGTGTCGTGGACAGCCCTCCCATAC 690
 |||||||
 OY 1405 CGTTACCATTCAGGGAATCTCTCCAAAGAGAGAGACTGTATACAG 1453
 |||||||
 Db 691 CGTTACCATTCAGGGAATCTCTCCAAAGAGAGAGACTGTATACAG 739
 |||||||
 RESULT 10
 AC119462/c 193189 bp DNA linear HTG 18-JUL-2002
 LOCUS Rattus norvegicus clone CH230-28N16, *** SEQUENCING IN PROGRESS
 DEFINITION *** 79 unordered pieces.
 AC119462
 AC119462.3 GI:21746590
 VERSION HTG: HTGS_PHASE1.
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 193189)
 Muzny,D.M., Adams,C., Adio-Odola,B., Ali-oshman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbada,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dihn,K.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotlo,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Loraado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mel,G., Mettaker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabdat,K., Morgan,N., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkenko,S., Ogun,M., Okunodu,G.,
 Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherier,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
 Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameris,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Welnsstock,G. and Gibbs,R.
 Direct Submission

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Unpublished
 2 (bases 1 to 193189)
 Worley,K.C.
 Direct Submission
 Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193189)
 Worley,K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20387269.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GURK
 Center clone name: CH230-28N16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 126440 bases at least Q40
 Consensus quality: 140831 bases at least Q30
 Consensus quality: 148852 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 79 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 1554: contig of 1554 bp in length
 * 1555
 * 1655: gap of unknown length
 * 1656
 * 2715: contig of 1061 bp in length
 * 2716
 * 2815: gap of unknown length
 * 2816
 * 3899: contig of 1084 bp in length
 * 3900
 * 3999: gap of unknown length
 * 4000
 * 5367: contig of 1368 bp in length
 * 5368
 * 5467: gap of unknown length
 * 5468
 * 6690: contig of 1223 bp in length
 * 6691
 * 6790: gap of unknown length
 * 6791
 * 8012: contig of 1222 bp in length
 * 8013
 * 8112: gap of unknown length
 * 8113
 * 9566: contig of 1454 bp in length
 * 9567
 * 9666: gap of unknown length
 * 10914: contig of 1248 bp in length
 * 10915
 * 11014: gap of unknown length
 * 11015
 * 12395: contig of 1381 bp in length
 * 12396
 * 12495: gap of unknown length
 * 14050: contig of 1555 bp in length
 * 14051
 * 14150: gap of unknown length
 * 14151
 * 15648: contig of 1498 bp in length
 * 15649
 * 15748: gap of unknown length
 * 15749
 * 16994: contig of 1246 bp in length
 * 16995
 * 17094: gap of unknown length
 * 17095
 * 18275: contig of 1180 bp in length
 * 18276
 * 18375: gap of unknown length
 * 18376
 * 20425: contig of 2052 bp in length
 * 20427
 * 20526: gap of unknown length
 * 22026: contig of 1500 bp in length
 * 22027
 * 22126: gap of unknown length
 * 23257: contig of 1131 bp in length
 * 23258
 * 23357: gap of unknown length
 * 23358
 * 24652: contig of 1295 bp in length

Consensus quality: 109854 bases at least Q40
 Consensus quality: 118111 bases at least Q30
 Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This is a 'working draft/sequence' It currently:

* consists of 55 contigs. The "true order" of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1547:	contig of 1547 bp in length
*	1548	1647:	gap of unknown length
*	1648	3064:	contig of 1417 bp in length
*	3065	3164:	gap of unknown length
*	3165	4885:	contig of 1721 bp in length
*	4886	4985:	gap of unknown length
*	4986	6000:	contig of 1015 bp in length
*	6001	6100:	gap of unknown length
*	6101	7171:	contig of 1071 bp in length
*	7172	7271:	gap of unknown length
*	7272	8794:	contig of 1523 bp in length
*	8795	8894:	gap of unknown length
*	8895	10368:	contig of 1474 bp in length
*	10369	10468:	gap of unknown length
*	10469	11906:	contig of 1438 bp in length
*	11907	12006:	gap of unknown length
*	12007	13031:	contig of 1025 bp in length
*	13032	13131:	gap of unknown length
*	13132	14575:	contig of 1444 bp in length
*	14576	14675:	gap of unknown length
*	14676	15599:	contig of 1024 bp in length
*	15599	15799:	gap of unknown length
*	15800	16951:	contig of 1152 bp in length
*	16952	17051:	gap of unknown length
*	17052	18695:	contig of 1644 bp in length
*	18696	18795:	gap of unknown length
*	18796	20470:	contig of 1675 bp in length
*	20471	20570:	gap of unknown length
*	20571	22418:	contig of 1848 bp in length
*	22419	22518:	gap of unknown length
*	22519	24464:	contig of 1846 bp in length
*	24365	24464:	gap of unknown length
*	24465	25586:	contig of 1122 bp in length
*	25587	27856:	gap of unknown length
*	27857	27959:	contig of 2173 bp in length
*	27960	29649:	gap of unknown length
*	29650	29749:	gap of unknown length
*	29750	31331:	contig of 1362 bp in length
*	31332	31431:	gap of unknown length
*	31432	33606:	contig of 2175 bp in length
*	33607	33706:	gap of unknown length
*	33707	35179:	contig of 1473 bp in length
*	35180	35797:	gap of unknown length
*	35280	37193:	contig of 1914 bp in length
*	37194	37993:	gap of unknown length
*	37294	38382:	contig of 1089 bp in length
*	38383	38482:	gap of unknown length
*	38483	40759:	contig of 2277 bp in length
*	40760	40859:	gap of unknown length
*	40860	42763:	contig of 1904 bp in length
*	42764	42863:	gap of unknown length
*	42864	45959:	contig of 3136 bp in length
*	46000	46099:	gap of unknown length
*	46100	49139:	contig of 3040 bp in length
*	49140	49239:	gap of unknown length
*	49240	51762:	contig of 2523 bp in length
*	51763	51862:	gap of unknown length
*	51863	54907:	contig of 3045 bp in length

*	54808	55007:	gap of unknown length
*	55008	57214:	contig of 2207 bp in length
*	57215	57214:	gap of unknown length
*	57315	60299:	contig of 2985 bp in length
*	60300	60399:	gap of unknown length
*	60400	62940:	contig of 2541 bp in length
*	62941	63040:	gap of unknown length
*	63041	65074:	contig of 2034 bp in length
*	65075	65174:	gap of unknown length
*	65175	68290:	contig of 3116 bp in length
*	68291	70436:	gap of unknown length
*	70436	70496:	contig of 2106 bp in length
*	70497	70596:	gap of unknown length
*	70597	73863:	contig of 3267 bp in length
*	73864	73863:	gap of unknown length
*	73864	76872:	gap of 2909 bp in length
*	76873	76872:	gap of unknown length
*	76913	79215:	contig of 2243 bp in length
*	79216	79315:	gap of unknown length
*	79316	83513:	contig of 4198 bp in length
*	83514	83613:	gap of unknown length
*	83614	87309:	contig of 3656 bp in length
*	87310	87409:	gap of unknown length
*	87410	92333:	contig of 4944 bp in length
*	92354	92453:	gap of unknown length
*	92454	95901:	contig of 3448 bp in length
*	95902	96001:	gap of unknown length
*	96002	100934:	contig of 4933 bp in length
*	100935	101034:	gap of unknown length
*	101035	104875:	contig of 3841 bp in length
*	104876	104975:	gap of unknown length
*	104976	109470:	contig of 4455 bp in length
*	109471	109570:	gap of unknown length
*	109571	117676:	contig of 8106 bp in length
*	117677	117776:	gap of unknown length
*	117777	124312:	contig of 6556 bp in length
*	124313	124412:	gap of unknown length
*	124413	128938:	contig of 4526 bp in length
*	128939	129038:	gap of unknown length
*	129039	135586:	contig of 6548 bp in length
*	135587	135686:	gap of unknown length
*	135687	143121:	contig of 7455 bp in length
*	143122	143231:	gap of unknown length
*	143232	149127:	contig of 5906 bp in length
*	149128	149227:	gap of unknown length
*	149228	156613:	contig of 7286 bp in length
*	156614	156613:	gap of unknown length

Query Match	11.9%;	Score 520;	DB 2;	Length 174346;
Best Local Similarity	93.9%;	Pred. No. 3e-88;		
Matches 541; Conservative	0;	Mismatches 35;	Indels 0;	Gaps 0;

OY	284	TTCCCGGCGCGCTGCGCGGCAAGAGCGCGGGCTTGTACTCCGGCGGGGGCGCC	343
Db	87286	TTCTACGGCTCTCGCGCGGCAAGGGCGGGGTGTATTACTCCGAGCGGGGCGCC	87227
OY	344	CGGCGACAGAGAGAGAGCTGGGGCGGTGTGTGCTCTTGCCCTGTCCGGCCCGGCAC	403
Db	87226	CGGCGACAGAGAGAGAGCTGGGGCGGTGTGTGCTCTTGCCCTGTCCGGCCCGGCAC	87167
OY	404	CAACAGCGCGCATTTGGGGGGAGCGCTGCCGACGCGAGCGCGCTGCAGCCCGCGCC	463
Db	87166	CAACAGCGCGCATTCGCGGGGAGCCAGTGCAGCGCGCGCTGCAGCCCGCGCC	87107
OY	464	CAAGTATCAGGCGGTGTGCCATTCTAGACGGGCTCTCTGTGGCGGCGCAAAAGGCC	523
Db	87106	CAAGTATCAGGCGGTGTGCCATTCTAGACGGGCTCTCTGTGGCGGCGCAAAAGGCC	87047
OY	524	TACGCCCTGGGCTGGGAGCAAGGTGGGGGGGCTCCCGCGTCGACACGGCTTCGGAGCC	583
Db	87046	TACGCCCTGGGCTGGGAGCAAGGTGGGGGGGCTCCCGCGTCGACACGGCTTCGGAGCC	86987
OY	584	GGCGGAGCCCGACCACTACTCTGTGCCGGGCGGCACCGCTTGCAGCCCGCACCGC	643

[illegible]

mus norvegicus, *Carnivora*: Chordata: Craniata: Vertebrata: Euteleostomi: *Mammalia*: Eutheria: Rodentia: Sciurognathii: Muridae; Murinae; Mus.

(bases 1 to 174346)

Abrahamson, S., Adams, C., Alder, J., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Aralbechi, V., Ayoyagi, R., Ayodeji, M., Baca, E., Baden, H., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Blair, J., Blankenburg, K., Blych, P., Brown, M., Buhan, N., Buch, P., Butrell, K., Calderon, E., Cadenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chen, Y., Chavez, J., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cockeand, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Gibbs, M. L., Davis, S., Day-Carroll, L., De Anda, C., Dedrich, D., Deppe, O., Dennis, S., Detamco, C., Ding, Y., Din, H., Diya, K., Dugan-Rocha, S., Dunn, A., Dudbin, K., Diva, B., Davies, K., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., George, C.M., Gablis, A., Gante, R., Garcia, A., Garner, T., Garza, M., Georgis, E., Geier, K., Gill, R., Grady, M., Guerra, M., Guervara, W., Harline, P., Healand, W., Hamill, C., Hamilton, C., Hamilton, K., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Ilms, B., Howells, S., Huliy, S., Hume, J., Idelbird, D., Jackson, A., Johnson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Joseph, S., Kelly, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J., Mensuhewa, L., Louisged, H., Lozano, R.J., Lu, X., Ma, J., Mahindratne, M., Mahmood, M., Malloy, K., Mangum, A., Mesherwati, M., Mahindratne, M., Martin, R., Martinez, E., Nguyen, B., Mapua, P., Martin, K., Martin, R., Mattheus, E., Minjaly, S., McLeod, M., Monelli, T., Meeneen, E., Milosavjevic, A., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Mustafa, M., Murphy, M., Naif, L., Nankervits, C., Neal, D., Nguyen, N., Norris, S., Nwokeneme, O., Okunuga, G., Orpanaagoun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plannoch, C., Plopper, F., Polndexter, A., Povovic, P., Pitman, E., Pu, L.-L., Puzo, M., Quitor, J., Rachlin, E., Regler, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Ritter, M., Richards, S., Riggs, F., Rives, C., Rodney, T., Rojas, A., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Shatsman, S., Shen, H., Shetty, J., Shwartsbeyn, A., Sitter, C.D., Smajz, D., Speedy, A., Sodergren, E., Soresle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Taylor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Trejos, Z., Usmani, K., Valias, R., Vera, V., Villaseña, D., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Wei, X., White, F., Williams, G., Willison, R., Wleczyk, R., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, X., Yeon, L., Yeon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zuo, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., L.R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Project Submission
Genome Sequencing Consortium.
(bases 1 to 174346)
Submitted (21-AGO-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc.help@bcm.tmc.edu
Project Information
Center project name: GCOU
Center clone name: CH320-278A19
Summary Statistics
Sequencing vector: plasmid,
Chemistry: dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1547: contig of 1547 bp in length
1548 1647: gap of unknown length
1648 3064: contig of 1417 bp in length
3065 3164: gap of unknown length
3165 4885: contig of 1721 bp in length
4886 6000: gap of unknown length
6001 6100: gap of unknown length
6101 7171: contig of 1071 bp in length
7172 7272: gap of unknown length
7273 8794: contig of 1523 bp in length
8795 8894: gap of unknown length
8895 10368: contig of 1474 bp in length
10369 10468: gap of unknown length
10469 11906: contig of 1438 bp in length
11907 12006: gap of unknown length
12007 13031: contig of 1025 bp in length
13032 13131: gap of unknown length
13132 14575: contig of 1444 bp in length
14576 14675: gap of unknown length
14676 15799: contig of 1024 bp in length
15700 15799: gap of unknown length
15800 16951: contig of 1152 bp in length
16952 17051: gap of unknown length
17052 18695: contig of 1644 bp in length
18696 18795: gap of unknown length
18796 20470: contig of 1675 bp in length
20471 20570: gap of unknown length
20571 22418: contig of 1848 bp in length
22419 22518: gap of unknown length
22519 24364: contig of 1846 bp in length
24365 24464: gap of unknown length
24465 25586: contig of 1122 bp in length
25587 27859: contig of 2173 bp in length
27860 27959: gap of unknown length
27960 29649: contig of 1690 bp in length
29650 29749: gap of unknown length
29750 31331: contig of 1582 bp in length
31332 31431: gap of unknown length
31432 33606: contig of 2175 bp in length
33607 33706: gap of unknown length
33707 35179: contig of 1473 bp in length
35180 35279: gap of unknown length
35280 37193: contig of 1914 bp in length
37194 37293: gap of unknown length
37294 38382: contig of 1089 bp in length
38383 38482: gap of unknown length
38483 40758: contig of 2277 bp in length
40759 40859: gap of unknown length
40860 42763: contig of 1904 bp in length
42764 42863: gap of unknown length
42864 45999: contig of 3136 bp in length
46000 46099: gap of unknown length
46100 49138: contig of 3040 bp in length
49139 49239: gap of unknown length
49240 51762: contig of 2523 bp in length
51763 51862: gap of unknown length
51863 54907: contig of 3045 bp in length
54908 55007: gap of unknown length
55008 57214: contig of 2207 bp in length
57215 57314: gap of unknown length
57315 60299: contig of 2985 bp in length
60300 60399: gap of unknown length
60400 62940: contig of 2541 bp in length
62941 63040: gap of unknown length
63041 65074: contig of 2034 bp in length

65075 65174: gap of unknown length
65175 68290: contig of 3116 bp in length
68291 68390: gap of unknown length
68391 70496: contig of 2106 bp in length
70497 70596: gap of unknown length
70597 73863: contig of 3267 bp in length
73864 73963: gap of unknown length
73964 76872: contig of 2909 bp in length
76873 76972: gap of unknown length
76973 79215: contig of 2243 bp in length
79216 79316: gap of unknown length
79317 83513: contig of 4198 bp in length
83514 83613: gap of unknown length
83614 87309: contig of 3696 bp in length
87310 87409: gap of unknown length
87410 92353: contig of 4944 bp in length
92354 92453: gap of unknown length
92454 95901: contig of 3448 bp in length
95902 96001: gap of unknown length
96002 100934: contig of 4933 bp in length
100935 101035: gap of unknown length
101036 104875: contig of 3841 bp in length
104876 104975: gap of unknown length
104976 109470: contig of 4495 bp in length
109471 109570: gap of unknown length
109571 117676: contig of 8106 bp in length
117677 117776: gap of unknown length
117777 124312: contig of 6536 bp in length
124313 124413: gap of unknown length
124413 128938: contig of 4526 bp in length
128939 129038: gap of unknown length
129039 135586: contig of 6548 bp in length
135587 135686: gap of unknown length
135687 143121: contig of 7435 bp in length
143122 143221: gap of unknown length
143222 149127: contig of 5906 bp in length
149128 149227: gap of unknown length
149228 156513: contig of 7286 bp in length
156514 156613: gap of unknown length

Query Match 11.9%; Score 520.6; DB 2; Length 174346;
Best Local Similarity 70.4%; Pred. No. 2.3e-88;
Matches 940; Conservative 0; Mismatches 339; Indels 56; Gaps 22;

QY 3022 CAAGACCTGCTGCTCAATATATAGGAATGCTTCTTACTTCAGACATATATG 3081
||
DB 95899 CAGNN 95958
DB 3082 TGAACAACAGTAGGGGCTATATCTCTAGAGGTAGGGCTTTATCTTAAGAGAA 3141
DB 95959 NNN 3142
QY 3142 TGTCCCGAGTTATATAGCACTTTAGAGGAGCAAGGATATGAGGTGTGGCTGG 3201
DB 96019 TGTCCCGGGA-TATTTACAC-TTTAGAGGAGAAAGCAAGTATGTAGGTGTGG 96076
DB 3202 CCCATGAGTGCAGAC---ATTCGAGAAATGGCATTGCCACGTGGAAGAGCAAGATTTC 96134
QY 96134 TCA-GGGCCGCCACACGCTAAACCTTTT-TGAGAGTGT-ATCTGTGCTCCCGAGGTTTG 96190
DB 3262 TCAAGGGGCTCCACACGTCTAAGTTTGTGTGAGATGTGATGTGCTTCCGTGATTG 3321
QY 3322 ACTTTAAAGCAATTTATCTGGAGACAACTAGTATCTTGTGATGATGTGCTGCTT 3381
DB 96191 ACTTTGAAGGAATTTATCCAGCAGTACACTTGTGTTCTGGATATCTTGCTGCTTT 96250
QY 3382 ATTTCCTCTTTGT 3441
DB 96251 AATTTCCTT-----TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 96303
QY 3442 ATCTGCTTGAAGAGAGTGAGGCTCTCTATTAAGGAACTGCTGTAACTTCAATTCACAGCAAG 3501

```

source
1. 615
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/notes="EMBL/GenBank Accession No. AF102093"
BASE COUNT      84 a      235 c      224 g      71 t      1 others
ORIGIN

Query Match      12.6%; Score 551.6; DB 6; Length 615;
Best Local Similarity 94.6%; Pred. No. 1.3e-94;
Matches 582; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

OY 112 GCGGCCCTGCGCGCGCGCAATCTGAGCAGCAGCTGAGTACGAGCGGCGTGGCGCTGG 171
    |||||
    1 CGGCGCCCTGCGCGCGCGCAACCCCGAGCGAGCTGAGTACGAGCGGCGTGGCGCTGG 60
OY 172 GCGGCCCGGAGAGCAGAGCTGGGCGCGCGAGAGCCCACTTCTCCCGCGCGCAAGC 231
    |||||
    61 GCGGCCCGGAGAGCAGAGCTGGGCGCGCGAGAGCCCACTTCTCCCGCGCGCAAGC 120
OY 232 TCAGAGAGCGGCGCGCGCGCGCTGCGCTCTCCAGAGGCGGAGCCCGCGCTTCCCGG 291
    |||||
    121 TCAGAGAGCGGCGCGCGCGCGCTGCGCTCTACCCAGAGGCGGAGCCCGCGCTTCCAG 180
OY 222 CCGGCT---GCGCGCGCGAGGCGCGCGCTTGTACTCCCGCGCGCGCGCGCGCGCG 348
    |||||
    181 CTGCGTGGCGCGCGCGCGAGGCGCGGCTTGTACTCCCGCGCGCGCGCGCGCGCG 240
OY 349 AGCAGAGAGAGAGCTGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 408
    |||||
    241 AGCAGAGAGAGAGCTGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
OY 409 AAGCCGCGATTTGGGCGGAGAGCTGCGCAGCGCGAGCGCGCTGCGCAGCCCGCGCGCAAGT 468
    |||||
    301 AAGCCGCGATTTGGGCGGAGAGCTGCGCAGCGCGAGCGCGCTGCGCAGCCCGCGCGCAAGT 360
OY 469 ATCAGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
    |||||
    361 ATCAGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 529 CTTGGGCTGGGAGACAAGGCTGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588
    |||||
    421 CTTGGGCTGGGAGACAAGGCTGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
OY 589 GACCCCGCACACTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648
    |||||
    481 GACCCCGCACACTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
OY 649 CCCTGGCGCGCGAGCGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708
    |||||
    541 CCCTGGCGCGCGAGCGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
OY 709 GCGGCTCGGAGAGCT 723
    |||||
    601 GCNCGCTGGGAGAGCT 615

```

```

RESULT 6
AC131356
LOCUS
DEFINITION
AC131356 174346 bp DNA linear HTG 21-AUG-2002
Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS
ACCESSION
AC131356
VERSION
AC131356.1 GI:22380622
KEYWORDS
HTG; HTGS; PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 174346)
Munzy, D., Marie, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

```

TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bissal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Buttrill, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, X., Dinh, H., Diver, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., George, J., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Guarnate, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hughes, M., Hollins, B., Howell, S., Huix, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbawa, L., Louised, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Maimoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedeme, O., Okonou, G., Olarinmusa, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., Puzos, M., Quatro, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shwartsbeyn, A., Sisson, I., Sitter, C., D., Smajs, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Stelmach, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Walidron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 174346)
Rat Genome Sequencing Consortium.
Submitted (21-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: G00J
Center clone name: CH230-278A19
-----Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 1000 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 109854 bases at least Q40
Consensus quality: 118111 bases at least Q30
Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces

Query Match	Best Local Similarity	Score	DB 2:	Length
Matches 799;	Conservative 0;	Mismatches 46;	Indels 3;	Gaps 1;
15	CCGACACCTGC	17.3%: 758.4;	DB 2:	193189;
Db	160907	CCGACACCTGC	94.2%: Pred. No. 1.3e-133;	
75	ACTATGACATGAGTATCGCGGTGTTC	94.2%: Pred. No. 1.3e-133;		
Db	160967	ACTATGACATGAGTATCGCGGTGTTC		

OY		135	AAAGCAGCGACTGAGATACGAGGGGGCTGGGGGTGGGGGGCCCGAGAGACGACTGGG	194
Dd		161027	GAGCAGCGACTGAGTACGAGGGGGCTGGGGGTGGGGGGCCCGAGAGACGACTGGG	161086
OY		195	GCAGCGCAAGGCCACTTCCTCCCGGGCACCCCTAAGCTCAAGAGCCGGGGCCCCCGCTG	254
Dd		161087	GCAGCGCAAGGCCACTTCCTCCCGGGCATCGTAAGCTCAAGAGCCGGGGCCCCCGCTG	161146
OY		255	GCCTCTCCCAAGGGGGGAGCCCGCGCTTCCCAGCCGCGT---GCGGGCAAGGGC	311
Dd		161147	GCCTCTCCCAAGGGGGGAGCCCGCGCTTCCCAGCCGCGT---GCGGGCAAGGGC	161206
OY		312	CGGGGCTTGTTACTCCCGGGGGGGGGGGGGGGGGGGGGGACAGGAAGAAGAGTGGGGCGGT	371
Dd		161207	CGGGGCTTGTTACTCCCGGGGGGGGGGGGGGGGGGGGGGACAGGAAGAAGAGTGGGGCGGT	161266
OY		372	TGCGTGCCTTGCCTCTGTCCGGCCCCCGGGCACCAAGCAAGCGGCATTGGGGGGGAGCCT	431
Dd		161267	TGCGTGCCTTGCCTCTGTCCGGCCCCCGGGGTACCAAGCAAGCGGCATTGGGGGGGAGCCA	161326
OY		432	GCCGACGCGGAGACCCTGCTGCAGCCCGCGGCCCAAGTATCAGGCGGTGCTGCCATTGAG	491
Dd		161327	GTCGAGCGCGGGGCTGCGCTGCAGCCCGCGGCCCAAGTATCAGGCGGTGCTGCCATTGAG	161386
OY		492	ACGGGCTCTCGTGGGGGGGGGGGCAAGAGCCCTAGGCCCCGAGGCTGGGGGACAAGAGGTGGG	551
Dd		161387	ACGGGCTCTCTGTTGGGGGGGGGGGCAAGAGCCCTAGGCCCCGAGGCTGGGGGACAAGAGGTGGG	161446
OY		552	GCAGGCTCCCGCGCTGCCACCGCTCGAGCCCGGGGGGACCCCAACACTACTCTGCCCC	611
Dd		161447	GCAGGCTCCCGCGCTGCCACCGCTCGAGCCCGGGGGGACCCCAACACTACTCTGCCCC	161506
OY		612	GGGGCGCCACCCCTCGCGCCCGCCACCGCGGGGAGCCCTGGGGGGGCGAGGGGAGAG	671
Dd		161507	GGGGCGCCACCCCTCGCGCCCGCCACCGCATGTGGGGAGCCCTGGGGGGGCGAGGGGAGAG	161566
OY		672	TGGAAGATATGAGAGAAAGAGCCCTTCGGGGGTGGTGGCGGCTCGGAGACCTTCAATCAG	731
Dd		161567	TGGAAGATATATGAGAGAAAGAGCCCTTCGGGGGTGGGGGGGTGGGGAGACCTTCAATCAG	161626
OY		732	GCCGCTCGCTCAAAACAGATCTTGTGCTGCAATTGGACTATGCAGACGACGACGAG	791
Dd		161627	GCCGCTCGCTCTCAAAACAGATCTTGTGCTGCAATTGGAGCTATGCAGACGACGACGAG	161686
OY		792	CAGCTGCAGGGCCAAAGAAAAGAGATPCGAGGAGCTGAAGTCAAGAGAGAGACAGCGCTT	851
Dd		161687	CAGTTGCAGGCCAAAGAAAAGAGATPCGAGGAGCTGAAGTCCGAGAGAGATACGGTACCG	161746
OY		852	GCTCGGAT	859
Dd		161747	GAGGGGT	161754
RESULT 5				
LOCUS	AX401049		615 bp	DNA linear PAT 06-JUN-2002
DEFINITION	Sequence 725 from Patent WO0210453.			
ACCESSION	AX401049			
VERSION	AX401049.1			GI:21337229
KEYWORDS				
SOURCE				
ORGANISM	Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and Elashoff,M.R.			
AUTHORS	Molecular toxicology modeling			
TITLE	Patent: WO 0210453-A 725 07-FEB-2002;			
JOURNAL	Gene Logic, Inc. (US)			
FEATURES	Location/Qualifiers			

Oy 4340 TAA 4342
 Db 50725 TAA 50723
 RESULT 4
 AC119462
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-28N16, *** SEQUENCING IN PROGRESS
 *** 79 unordered pieces.
 AC119462 193189 bp DNA linear HTG 18-JUL-2002
 AC119462.3 GI:21746590
 HTG: HTGS-PHASE1.
 VERSION HTG: HTGS-PHASE1.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 193189)
 Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
 Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delany,K.R., Delgado,O., Denn,A.L., Ding,Y., Dint,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Krtolovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Mathew,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
 Soderren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umanal,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wlarczyk,R., Woodem,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 193189)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193189)
 REFERENCE Worley,K.C.
 AUTHORS Direct Submission
 TITLE Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 JOURNAL Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 14, 2002 this sequence version replaced gi:20387269.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hpsc-help@bcm.tmc.edu

 Project Information
 Center project name: GURK
 Center clone name: CH230-28N16

 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye 1008 of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 126440 bases at least Q40
 Consensus quality: 140831 bases at least Q30
 Consensus quality: 146852 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 79 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1554: contig of 1554 bp in length
 * 1555 1654: gap of unknown length
 * 1655 2715: contig of 1061 bp in length
 * 2716 2815: gap of unknown length
 * 2816 3899: contig of 1084 bp in length
 * 3900 3999: gap of unknown length
 * 4000 5367: contig of 1368 bp in length
 * 5368 5467: gap of unknown length
 * 5468 6690: contig of 1223 bp in length
 * 6691 6790: gap of unknown length
 * 6791 8012: contig of 1222 bp in length
 * 8013 8112: gap of unknown length
 * 8113 9566: contig of 1454 bp in length
 * 9567 9666: gap of unknown length
 * 9667 10914: contig of 1248 bp in length
 * 10915 11014: gap of unknown length
 * 11015 12395: contig of 1381 bp in length
 * 12396 12495: gap of unknown length
 * 12496 14050: contig of 1555 bp in length
 * 14051 14150: gap of unknown length
 * 14151 15648: contig of 1498 bp in length
 * 15649 15748: gap of unknown length
 * 15749 16994: contig of 1246 bp in length
 * 16995 17094: gap of unknown length
 * 17095 18274: contig of 1180 bp in length
 * 18275 18374: gap of unknown length
 * 18375 20426: contig of 2052 bp in length
 * 20427 20526: gap of unknown length
 * 20527 22126: contig of 1500 bp in length
 * 22127 23257: contig of 1131 bp in length
 * 23258 23357: gap of unknown length
 * 23358 24652: contig of 1295 bp in length
 * 24653 24752: gap of unknown length
 * 24753 26187: contig of 1435 bp in length
 * 26188 26287: gap of unknown length
 * 26288 27567: contig of 1280 bp in length
 * 27568 27667: gap of unknown length
 * 27668 28980: contig of 1313 bp in length
 * 28981 29080: gap of unknown length
 * 29081 30136: contig of 1056 bp in length
 * 30137 30236: gap of unknown length
 * 30237 31632: contig of 1396 bp in length
 * 31633 31732: gap of unknown length
 * 31733 33270: contig of 1538 bp in length
 * 33271 33370: gap of unknown length

QY	3464	TCCTAAAGGGGAACCTGCGTGAACCTGATTCGATTCGACACGAAGATGAGAGAAATAGGACTT	3523
Db	1441	TCCTAAAGGGGAACCTGCTGTGAACCTGATTCGATTCGACACGAAGATGAGAGAAATAGGACTT	1500
QY	3524	AATTCACACTAGGGGCTCTCATCTCACACACTTAAAGAGAGAGATTTCTAGAAAAACTGGGCC	3583
Db	1501	AATTCACACTAGGGGCTCTCATCTCACACACTTAAAGAGAGAGATTTCTAGAAAAACTGGGCC	1560
QY	3584	AGATTTTCTTGTGTTCTCCATCATTTTAAATGAGGAGAGCTGTTCAGTTTTCTTACTCTTAC	3643
Db	1561	AGATTTTCTTGTGTTCTCCATCATTTTAAATGAGGAGAGCTGTTCAGTTTTCTTACTCTTAC	1620
QY	3644	CTATGATATATTTCTTGCTGAACGCTGTCACAAAAGAAAAAGAACCCAAATCACTGTCCTTG	3703
Db	1621	CTATGATATATTTCTTGCTGAACGCTGTCACAAAAGAAAAAGAACCCAAATCACTGTCCTTG	1680
QY	3704	ACTTTGTTCTTGTGATCCCTCAGTTTCTTGTGATTTGACAGTGTGCGGGTCTCTAAATT	3763
Db	1681	ACTTTGTTCTTGTGATCCCTCAGTTTCTTGTGATTTGACAGTGTGCGGGTCTCTAAATT	1740
QY	3764	TGGGTATGAGTTACCAAAATTTAACCATTTGTTGTGCCCCACCCAGGGGACCTCCACGT	3823
Db	1741	TGGGTATGAGTTACCAAAATTTAACCATTTGTTGTGCCCCACCCAGGGGACCTCCACGT	1800
QY	3824	TTCTGACTTGAAGTAGACTGAGAAATCCACGAGGTCTATCTGGCCAGATTTAACTAG	3883
Db	1801	TTCTGACTTGAAGTAGACTGAGAAATCCACGAGGTCTATCTGGCCAGATTTAACTAG	1860
QY	3884	ATTCTATTTCCTGGTCTCCCTCCCTCCGTCGAGACCTCTTATTTATTTGCCCTCTCT	3943
Db	1861	ATTCTATTTCCTGGTCTCCCTCCCTCCGTCGAGACCTCTTATTTATTTGCCCTCTCT	1920
QY	3944	AGGTTAAATTCCTCTTGTGATTTGCACTTTGTTGAGAGAGAGGTTGAGACAGTAGATTAGCATA	4003
Db	1921	AGGTTAAATTCCTCTTGTGATTTGCACTTTGTTGAGAGAGAGGTTGAGACAGTAGATTAGCATA	1980
QY	4004	GTTCACAGTGCAAAATTTACAGTGTGTTAGAGATGTGGGGGAAAAATTAGCTTATTTTTC	4063
Db	1981	GTTCACAGTGCAAAATTTACAGTGTGTTAGAGATGTGGGGGAAAAATTAGCTTATTTTTC	2040
QY	4064	CTACATGGGATACAAACACTGTGAATTCATCTTAACTGAGAGGCCCTGCACATTCCTCTAA	4123
Db	2041	CTACATGGGATACAAACACTGTGAATTCATCTTAACTGAGAGGCCCTGCACATTCCTCTAA	2100
QY	4124	AACATAGTGTGTTGTTTCTTTTAAACAAGTTTAAAGCTAGTGTATATATATTAATAAAAAA	4183
Db	2101	AACATAGTGTGTTGTTTCTTTTAAACAAGTTTAAAGCTAGTGTATATATATTAATAAAAAA	2160
QY	4184	TTGCTGTGCTGCTACCTCAGCTTGTTTATGCGCATTTGATATTTGTTGCTGCTGTGT	4243
Db	2161	TTGCTGTGCTGCTACCTCAGCTTGTTTATGCGCATTTGATATTTGTTGCTGCTGTGTGT	2220
QY	4244	AATTCATTAACCTTTTGATACCAATTTCTGATGTGTAAAAATGGTGTGCTGTGTAATATCTTA	4303
Db	2221	AATTCATTAACCTTTTGATACCAATTTCTGATGTGTAAAAATGGTGTGCTGTGTAATATCTTA	2280
QY	4304	TAAAGAGTTCAATTGT 4319	
Db	2281	TAAAGAGTTCAATTGT 2296	

AL590963/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AL590963	196724 bp	DNA	linear	ROD 05-APR-2002			
	Mouse DNA sequence from clone RP23-395E10 on chromosome 11, complete sequence.						
AL590963							
AL590963.11	GI:20068458						
HTG.							
house mouse.							
Mus musculus							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eultheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Lovell, J.	Direct Submission	Submitted (04-APR-2002)	Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Apr 7, 2002 this sequence version replaced gi:16304739.
				During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
				This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at: http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-395E10 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.choi1.org/bacpac/home.htm VECTOR: PBACe3.6.
FEATURES	SOURCE	location/Qualifiers		
		1..196724		
		/organism="Mus musculus"		
		/db_xref="taxon:10090"		
		/chromosome="11"		
		/clone="RP23-395E10"		
		/clone_11b="RPCI-23"		
BASE COUNT	49842 a	49408 c	48640 g	48834 t
ORIGIN				
Query Match	18.9%	Score 825;	DB 10;	Length 196724;
Best Local Similarity	69.7%;	Pred. No. 2.7e-146;		
Matches 1841;	Conservative	0;	Mismatches 540;	Indels 262; Gaps 43;
Oy	1832	GAATTGGAGCTACCTCGTGTGATGAGCGGATGCGATTTGGAGATCCGAGAAG	1891	
Db	53235	GAACTCTGAATACCTCGGTGGATGAGCGAAGCCGTTCGATGAGATCCAGAAG	5317	
Oy	1892	GCAAAACACCTCACCGAGCTGTAGAAATAGCTGTGCGCAAAACCTGTCTTCAGAT	1951	
Db	53175	GCAACACACTACCGGACCTGTAGAAATAGCTGTGCGC-GGTACCTGTCTTCAGAT	53111	
Oy	1952	AGTTGTAGCATGCATTC-CGAGAGTGGCAGAGACCTGTAAATGTGACCTTGTCTCA	2010	
Db	53116	AGTTGTAGCATGTCTGTTCGAGAAATAGCAGACACCTGTAAATGTGACCTGTCTCA	5305	
Oy	2011	CATATGTTATACTCGCTGATTAATACCTTTCATACCTCTTGA-CTTGTCTTC	2064	
Db	53056	TGCTGTATTCATTCACCTGGTAAAGCCCTTCACCTCCCTTAAATAGATTGTCTTC	5299	
Oy	2065	ATTACTCGATTTCACAA-----AAACTCTTCAATTCGGGCAATTTGAGTTA	2112	
Db	52996	TTTACTCGATTTCACAAACAAACAGAAACCTCTTCGCTGGCTGATTTGAAATTA	5293	
Oy	2113	TGAGAGGTGATTTGGATTTCTTTTCCCTTTTGGAAATGCGCTCAAGCTAAACCTA	2172	
Db	52936	---GGGCTACTGAGAGCGGCTTTTCCATTTTGGGAGATGAG--CTTGACCTGTCTTA	5288	
Oy	2173	TAGATGCGACATTCAGAAAGTTTCAGGGGCTGTCTTATACATTTGCCATGTAAAG	2232	
Db	52881	TAGG--GGAGACATTTGGGGGCTCCAGAGGCGCACATTGAACACACCTGCTCCATAAGGG	5282	
Oy	2233	GCTAAAGGCTCTCTTCATTTAGCATGTGAGAAATGACAGCCCTCTCTTTAGAGCT	2292	

OY 2252 TTAGACATGTGGAAGATGAGAGAGCCCTTCCTTTAGAGCTGCTGCATGGCCTCT 2311
DB 78303 TTAGCATGTGGAAGATGAGAGAGCCCTTCCTTTAGAGCTGCTGCATGGCCTCT 78382
OY 2312 CTCACCCGTGGTACACCCCTCTTATAGTGGATATAGTATTTTAAACCTTAATAAACA 2371
DB 78363 CTCACCCGTGGTACACCCCTCTTATAGTGGATATAGTATTTTAAACCTTAATAAACA 78422
OY 2372 AACAACTTCACCATGAGCTTTAGACCCAGAGAGATGACAGATGAAAGCCATGAGCA 2431
DB 78423 AACAACTTCACCATGAGCTTTAGACCCAGAGAGATGACAGATGAAAGCCATGAGCA 78482
OY 2432 GCCATCTTCACAGATGAGAGAGAGATGAGATGATTAATCTGCTGAAGAATA 2491
DB 78483 GCCATCTTCACAGATGAGAGAGATGAGATGATTAATCTGCTGCTGAAGAATA 78542
OY 2492 GTTGTTCATTTGAAACTATCTGTGATACAGTATGAGGAGAGGATGTTGGCTGTAT 2551
DB 78543 GTTGTTCATTTGAAACTATCTGTGATACAGTATGAGGAGAGGATGTTGGCTGTAT 78602
OY 2552 TATTTTTCAGTATATGATATACATTTCTTACGCTCAAAAACCAAAATCTTGGAAA 2611
DB 78603 TATTTTTCAGTATATGATATACATTTCTTACGCTCAAAAACCAAAATCTTGGAAA 78662
OY 2612 AGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCTGTAAACAAAGCACTAGTAT 2671
DB 78663 AGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCTGTAAACAAAGCACTAGTAT 78722
OY 2672 TTGGGATGGCATGCCAAAACCTGTATAATGTCTTGTATCAGTACATCTCTCAAGTAT 2731
DB 78723 TTGGGATGGCATGCCAAAACCTGTATAATGTCTTGTATCAGTACATCTCTCAAGTAT 78782
OY 2732 CCTTCATTTGGGCTTCATCTTTTAGCAGAACCTGTGGTGGGATGATGAGCTTGGGAG 2791
DB 78783 CCTTCATTTGGGCTTCATCTTTTAGCAGAACCTGTGGTGGGATGATGAGCTTGGGAG 78842
OY 2792 GGTAGGGGAGAGTGTGAAATAGTGTCTCTTGGCTGGCAATGTCTACATCTTGA 2851
DB 78843 GGTAGGGGAGAGTGTGAAATAGTGTCTCTTGGCTGGCAATGTCTACATCTTGA 78902
OY 2852 ACAAAACAGATGATCTAATGAGCTTCTCATTACCTTTGTAAAAATTTGTATGTGA 2911
DB 78903 ACAAAACAGATGATCTAATGAGCTTCTCATTACCTTTGTAAAAATTTGTATGTGA 78962
OY 2912 CCATCTGTGCTCCTCCCTCCCTCGTTTGTAAAAATTCAGAGATGAGCTCCAGGCCAC 2971
DB 78963 CCATCTGTGCTCCTCCCTCCCTCGTTTGTAAAAATTCAGAGATGAGCTCCAGGCCAC 79022
OY 2972 TTTGCTCAGTGTAAAGTCCCTATTAATCTGTAAGAGAAATAGAGCCAGACCTCT 3031
DB 79023 TTTGCTCAGTGTAAAGTCCCTATTAATCTGTAAGAGAAATAGAGCCAGACCTCT 79082
OY 3032 GGTCTCAAAATATATAGAAATGCTTCTTATAGTCTTACAGACTTGTGTGAAAAACAAG 3091
DB 79083 GGTCTCAAAATATATAGAAATGCTTCTTATAGTCTTACAGACTTGTGTGAAAAACAAG 79142
OY 3092 TAGGGGTCTAATCTCTAGAAAGTGGGGCTTTTATCTTAAAGAGATATGTCCCGAGA 3151
DB 79143 TAGGGGTCTAATCTCTAGAAAGTGGGGCTTTTATCTTAAAGAGATATGTCCCGAGA 79202
OY 3152 TTATTTAGCATTTTATAGAGAGAGCAAGGATATGAGGATGTGTGCTGGCCACAGCTG 3211
DB 79203 TTATTTAGCATTTTATAGAGAGAGCAAGGATATGAGGATGTGTGCTGGCCACAGCTG 79262
OY 3212 GAGCAGCAGAGAGAGATGGATTCACATTTGTGGGAGAGAGAAAAATCTCTAGGGGCT 3271
DB 79263 GAGCAGCAGAGAGAGATGGATTCACATTTGTGGGAGAGAGAAAAATCTCTAGGGGCT 79322
OY 3272 CCCACTGCTAAGTTTTGTGAGATGTGTGATCTGTCTCTGATTTGACTTTTAAG 3331
DB 79323 CCCACTGCTAAGTTTTGTGAGATGTGTGATCTGTCTCTGATTTGACTTTTAAG 79382

OY 3332 GAATTAATCTGGACACACATGATGATTAATCTTGATATGATCTTCTCTATTTCTCCT 3391
DB 79383 GAATTAATCTGGACACACATGATGATTAATCTTGATATGATCTTCTCTATTTCTCCT 79442
OY 3392 TTGT 3451
DB 79443 TTGT 79502
OY 3452 CAGAGTGGGCTCTCATTAAGGAGAACCTGTGTAACTTCATTTGACGAGCAAGATGTAGGA 3511
DB 79503 CAGAGTGGGCTCTCATTAAGGAGAACCTGTGTAACTTCATTTGACGAGCAAGATGTAGGA 79562
OY 3512 GAATATGAGACTTAATTCACATGAGGGCTCTCATTCACACCTTAAGAGAGATTTCTAG 3571
DB 79563 GAATATGAGACTTAATTCACATGAGGGCTCTCATTCACACCTTAAGAGAGATTTCTAG 79622
OY 3572 AAAAAGTGGCCAGATTTTCTTGTGTTCATCATTTTAATGTGGAGGCTGTTCAGTTT 3631
DB 79623 AAAAAGTGGCCAGATTTTCTTGTGTTCATCATTTTAATGTGGAGGCTGTTCAGTTT 79682
OY 3632 TCTTACTCTTACCTATGATATTTCTTCGTACGTGTCCAAAAAGAAAAAGCCCAAT 3691
DB 79683 TCTTACTCTTACCTATGATATTTCTTCGTACGTGTCCAAAAAGAAAAAGCCCAAT 79742
OY 3692 CAGTGTCTTGTACCTTGTCTTGTATGATCCCTCAGTCTCTCTTGTATTCACAGATGTCTG 3751
DB 79743 CAGTGTCTTGTACCTTGTCTTGTATGATCCCTCAGTCTCTCTTGTATTCACAGATGTCTG 79802
OY 3752 GGTTCCTAATTTTGGGATATGAGTTAGCAAAATTTAACCTTTGTGTGTGCTCCACAGG 3811
DB 79803 GGTTCCTAATTTTGGGATATGAGTTAGCAAAATTTAACCTTTGTGTGTGCTCCACAGG 79862
OY 3812 GGACTGCCAGTTTGTGACTGGAATGATAGAGAGAGAGATCCAGAGGTGTATGTGCC 3871
DB 79863 GGACTGCCAGTTTGTGACTGGAATGATAGAGAGAGAGATCCAGAGGTGTATGTGCC 79922
OY 3872 AGATTTAAGTATCTATTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931
DB 79923 AGATTTAAGTATCTATTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 79982
OY 3932 GTCCCTCTTCTAGTATTAATCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3991
DB 79983 GTCCCTCTTCTAGTATTAATCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 80042
OY 3992 TAGATTACCAAGTCCCAAGTGCAGAAATTTAAGATGTGAGTGGGGGAAAAATTAG 4051
DB 80043 TAGATTACCAAGTCCCAAGTGCAGAAATTTAAGATGTGAGTGGGGGAAAAATTAG 80102
OY 4052 TCTTATTTTCCCTACATGGGATACAGACGTGTGAATTCATCTTCAACTGAAGCCCTG 4111
DB 80103 TCTTATTTTCCCTACATGGGATACAGACGTGTGAATTCATCTTCAACTGAAGCCCTG 80162
OY 4112 CAGTCTCTTAAACATAGT 4171
DB 80163 CAGTCTCTTAAACATAGT 80222
OY 4172 AATTAATAAAATGCTGT 4231
DB 80223 AATTAATAAAATGCTGT 80282
OY 4232 TGTCTGTGTGTATTCATTAACCTTTGTATACATTTGTATGTGTAAATTTGTTGTCTT 4291
DB 80283 TGTCTGTGTGTATTCATTAACCTTTGTATACATTTGTATGTGTAAATTTGTTGTCTT 80342
OY 4292 GTAATATCTTATTAAGAGTTCATTTGTAATTAACATTTGTGTGTGTAA 4342
DB 80343 GTAATATCTTATTAAGAGTTCATTTGTAATTAACATTTGTGTGTGTAA 80393

RESULT 2
AK055378
LOCUS AK055378 2296 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ30816 f1s, clone FBRA2001571.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:18 ; Search time 10970 Seconds
(without alignments)
11598.675 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagcctcgcagcccccgcac.....aaaaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	57.4	180017	2	AC068669	AC068669 Homo sapi
2	2294.4	52.5	2296	9	AK055378	AK055378 Homo sapi
3	825	18.9	196724	10	AL590963	AL590963 Mouse DNA
4	758.4	17.3	193189	2	AC119462	AC119462 Rattus no
5	551.6	12.6	615	6	AX401049	AX401049 Sequence
6	520.6	11.9	174346	2	AC131356	AC131356 Rattus no
7	520	11.9	174346	2	AC131356	AC131356 Rattus no
8	510.2	11.7	542	6	AX368844	AX368844 Sequence
9	380.2	8.7	1433	9	HS4800235	AL049450 Homo sapi
10	371	8.5	193189	2	AC119462	AC119462 Rattus no
11	274.8	6.3	590	6	AX400340	AX400340 Sequence
12	242	5.5	207214	2	AC107643	AC107643 Mus muscu
13	240	5.5	240	11	G22752	G22752 human STS W
14	235.4	5.4	135060	9	AL354659	AL354659 Human DNA
15	235.4	5.4	142808	2	AL513172	AL513172 Homo sapi
16	235.4	5.4	165311	9	AC097061	AC097061 Homo sapi
17	213.2	4.9	177607	2	AC119484	AC119484 Rattus no
18	212	4.8	171318	2	AC127122	AC127122 Rattus no
19	211.2	4.8	201143	2	AC125268	AC125268 Mus muscu
20	166	3.8	207214	2	AC107643	AC107643 Mus muscu
21	150.2	3.4	144577	2	AC123367	AC123367 Rattus no
22	107.2	2.5	565	6	AX401061	AX401061 Sequence
23	92.6	2.1	72397	2	AC114443	AC114443 Rattus no
24	90	2.1	132727	2	AC120736	AC120736 Rattus no
25	87.4	2.0	175440	2	AC126197	AC126197 Rattus no
26	87.2	2.0	74721	2	AC096161	AC096161 Rattus no
27	86.6	2.0	991	11	PM128128	AL684455 Penicilliu
28	85.2	1.9	184402	2	AC127041	AC127041 Rattus no
29	84.2	1.9	885	11	PM7F8G	AL685189 Penicilliu
30	83.8	1.9	55802	2	AC103271	AC103271 Rattus no
31	83.8	1.9	150090	2	AC121750	AC121750 Rattus no
32	83.4	1.9	202509	2	AC129875	AC129875 Rattus no
33	83.2	1.9	1094	11	PM7G11B	AL685196 Penicilliu
34	83.2	1.9	145592	2	AC119341	AC119341 Rattus no
35	82.8	1.9	300695	2	AC079431	AC079431 Mus muscu
36	81.8	1.9	26085	2	AC096530	AC096530 Rattus no
37	81.6	1.9	61958	2	AC121467	AC121467 Rattus no
38	81.6	1.9	101282	2	AC121474	AC121474 Rattus no
39	81.6	1.9	195227	2	AC129795	AC129795 Rattus no
40	81.4	1.9	125020	9	AR429315	AR429315 Homo sapi
41	80.8	1.8	69217	2	AC129516	AC129516 Homo sapi
42	80.8	1.8	136014	2	AC126531	AC126531 Rattus no
43	80.8	1.8	174410	2	AC128158	AC128158 Rattus no
44	80.6	1.8	179608	2	AC128497	AC128497 Rattus no
45	80.4	1.8	1065	11	PM2B12B	AL684695 Penicilliu

ALIGNMENTS

RESULT 1
AC068669 180017 bp DNA linear HTG 16-FEB-2002
LOCUS Homo sapiens chromosome 17 clone RP11-749116 map 17, *** SEQUENCING
DEFINITION IN PROGRESS ***, 16 unordered pieces.
ACCESSION AC068669
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULTOP; HTGS_ACTIVEPIN.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180017)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-749116


```

QY 331 CCGGGGCGGGCCCCGGGACAGAGAGAGCTGGGGGCTGCGTCCCTTGCCCTGTC 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4208 CAGGCTCCGGACAGAGTCCGGCTCTCCGGCTCTCCGGCTCTCCGGCTCTCCGGCT 4149
QY 391 CGCCCCCGGGCCACCAAGCAAGCCGCAATGGGGGGGAGCCCTGCCGACCGGAGCGGCT 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4148 CCGGCGCCGACCGGACCGGCGGACCGGCGGCAAGAGCCCGCGCGCGGCGGCGGCGG 4089
QY 451 GCAGCCCCCGGGCCCAATGATCAGGCGGTGCT-GCCCATTCAGACGGGCTCTCTCGGCG 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4088 GCCCCCGGGAGCGGAGAGAGAGAGAGGCTCTCCGGCTCTCCGGCTCTCCGGAGCGGC 4029
QY 510 GCGGCGCAAGAGCTTACGCTTGGGCTGGGAGACAGGTTGGGGCGGCTTCCCGCTGCC 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4028 ACGGCGACCGGACGAGAGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAC 3969
QY 570 ACCGCTCGAGCCCGGGGAGACCCCGACACACTCTGCTGCCGCGGCGGCGGCGGCGG 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3968 TCGGGCGCGCTCGGAGACCCCGCTCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 3909
QY 630 CCGACCGCGCACGCGC 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3908 CCGAGGACGACGCGC 3894

```

RESULT 13

```

US-07-867-106-2/c
: Sequence 2, Application US/07867106
: Patent No. 5389526
: GENERAL INFORMATION:
: APPLICANT: Slade, Martin B
: APPLICANT: Slade, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
: STREET: One Liberty Place 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/867,106
: FILING DATE: 19920625
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU PJ 7187
: APPLICATION NUMBER: PCT/AU90/00530
: FILING DATE: 02-NOV-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Feeney, Joanne Longo
: REGISTRATION NUMBER: 35,134
: REFERENCE/DOCKET NUMBER: RICE-0002
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5852 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS

```

```

: LOCATION: 2378..5038
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2378..5038
US-07-867-106-2

```

```

Query Match 1.3%; Score 56.4; DB 1: Length 5852;
Best Local Similarity 53.1%; Pred. No. 0.0023;
Matches 120; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```

```

QY 4145 TTACAAAGTTTAACTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 4204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5818 TAAAGATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5759
QY 4205 CTTTGTTTATGGCCATTTATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5758 TTGTTATTTTATTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 5699
QY 4265 TTTCGATGTGTAATTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5698 TTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5639
QY 4325 AACTATTTGCTGTGTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5638 AATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5593

```

RESULT 14

```

US-09-103-840A-2/c
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentin Ver. 2.1
: LENGTH: 4403765
: SEQ ID NO 2
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Query Match 1.3%; Score 56; DB 4: Length 4403765;
Best Local Similarity 46.3%; Pred. No. 0.069;
Matches 304; Conservative 0; Mismatches 335; Indels 17; Gaps 3;

```

```

QY 9 GACCCCCCGACCTGCGCCCTTCCGCCACCCCTCTCCGCTGCGGTCGCCGGGCTGCTC 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841630 GCCCGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841571
QY 69 CGGACCACTATGACATGAGATCCGCGGTTGAAGCGGCGCGGCGGCGGCGGCGGCGGCG 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841570 CCGTTCCGATCAACCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841511
QY 129 AATCTGAGCAGCGACTGACATGACAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841510 CTACTATATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 841451
QY 189 CCGTGGGGGGGCGGAGCCCACTTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841450 CCGGCTCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 841391

```


OY	257	TTCCATCCAGGAGCGGAGCCCGCGGCTTTCCCGCGCGGCTCGCGCGGCAAGGCGCGGG	316
Db	92894	CGCCGCCAGACGCCCGGTCGCGCTCCCGCGCGCGCTTCCCGCGCGGTTCGGTGG	929003
OY	317	CTTGTTACTCCCGCGCGGCGGCGCCCGCGGCGAGAGAGAGAGCTGGGCGGTTTCGGT	376
Db	929004	CGGCTCCSCAACCCCGCGAGCGCGCGATCGCGGCGCGCATTCGAGAGGATCGG	929063
OY	377	GCCTTGGCCCTGTCCGCGCGCGGCGCACCAAGAACCGGCGATTGGGGGGAGCCTCCGC	436
Db	929064	CGCGGTCCTCCCATGGCGCGCTACAGCCGAAGCGCGCGCGCGGACCGCCGCGCG	929123
OY	437	AGCGGAGACCGGCTGCAGCCCGCGGCCAAATGATCAGCGGCTGCTCCCATTCAGACGG	496
Db	929124	CGCGGCGCGCGCC -- CAGCAGCGCGACGCCCAATCGCGCGCGCGCATGCGCCGGT	929181
OY	497	CTCTCTCGTGGCGGCGGCCAAGAGACCTTACGCCCTGGGCTGGGGACAGAGGTTGGCGGC	556
Db	929182	CTCGGTGGCGCGCGCGCCAGCGCCCGCGCGCGCGCGCGCGCCAGCGCCACAGACCGCGAA	929241
OY	557	CTCCCGCGCTGCACCGGCTCGGAGCCCGGCGGGAGCCCGCACATTACTCTGCGCGGCGC	616
Db	929242	CGCGCGCGGACCAAGCGCACCGGTCCCGCGGTGCGCGGCAACCGGTGCAGACCCACAC	929301
OY	617	GCCACCCCTCGCGGCCACCGCACCGCGCGGAGACCTTGGCGGCGACGAGGCGCATGTGAA	676
Db	929302	AAGCCCGCGCGCGCGGTAGCGCGCGCGCGCAACCGATGAAAGTGGGGGGCAACACGAA	929361
OY	677	G 677	
Db	929362	G 929362	

```

RESULT 10-537-1
US-09-105-537-1
: Sequence 1, Application US/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D. H.
: APPLICANT: Liu, H.
: APPLICANT: xue, Y.
: APPLICANT: zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and plkrmycycin
: FILE REFERENCE: 600.4380U1
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 15872
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	1.3%;	Score 57.6;	DB 4;	Length 15872;
Best Local Similarity	46.1%;	Pred. No. 0.0019;		
Matches 298;	Conservative	0;	Mismatches 344;	Indels 4;
			Gaps	3

QY	13	CCCCGACCTCGCCCTCTCCCAACCCCTCTCGCGCTGGTGCCCGGCGCTGCTCCGA	72
Db	11368	CCCCGTACCGCTCCCGTGAGCTCCCGGCCCGGACCGGGCGCTTCTCGCGCACTGCT	11427
QY	73	CCACTATGACCATGAGATCCGCGGTGTTCAAGCGGCGCGCCCTGCGCGCGCATTC	132
Db	11428	GCGGTCGC - CCACCGGACCTGCACACAGAGTGACACACCGCGCCGTGCGCGGTGC	11486
QY	133	CTGAGCAGCGCACTGAGCTACGACGCGGCTCGCGCTGGCGGGGCCCGAGAGCAACTTG	192
Db	11487	TGTGCGTGTCTCCGAGAGAGATGCGGCCCATCGCAGACGCGCCGTAACCGCGGGG	11546
QY	193	GGGCGCGCAAGCCCACTTCTCCCCCGGACCGCTAAGCTCAAGAGACCGGGGCCCGCC	252

Db	11547	TCCTGGGAGCGCTGCTCCCTGATGCAAGCTATGGAGGAGGAGCGGTGGAGGCTCGCGTGT	11606
OY	253	TGCGCTCTCTCCCAAGGCGGGAGGCCCGCGGCTTCCCGCGCGGCTCGCGCGGCMAAGGCC	312
Db	11607	GGTCCGCTCTCCCGCGCGCGCGCGTGCACCGCCCGACGGGAGAGCCCGTTCGCGCGGGG	11666
OY	313	GGGCGCTGTTACTGCCCGCGGGGAGCGGCGCCCGGAGCAGCAGAAAGACATGTGGGCGGTT	372
Db	11667	CCGCGCTGTGGGGGCGTGGGCGGGGTGGCCGCTTGGAGAGCCCGCACCGGTTGGGCGGTC	11726
OY	373	CGGTGCCCTTGCCCTGTCCGCGCGCCCGGACACCAAGCAAGCCGGCATTTGGGGGGAGCCTG	432
Db	11727	TCGTGGACACTGCCCGCGCTCCCGCGGTGCGGCGACACTGGGCGGCGCGCTGGAAAGGCTCG	11786
OY	433	CCGCAAGCC--GAGACCGGCTGACACCCCGGCGCCAGTATCAGAGCGGTCTGCCATTCA	490
Db	11787	CCGGTCTCCGAGAACCAATTCGCGTGGCGCGCTGTGCGGCAAGTTGGGGCGCGGCGCTTCACCA	11846
OY	491	GACGGGCTCTGTCGTGGCGCGCGGACCAAGAGCCTTAAGCCCTGGGCTGGGAGACAGGGTG	550
Db	11847	GGCTGGCGGCG--GACGGCGGCGGGCGGAGCGGCGACACCGCGTACCGGCGCGCGGCAAG	11905
OY	551	GGCGGCTTCCCGCGCTGCCACGCGCTGGACCCCGGCGGAGACCCCAACAATACTTCGCC	610
Db	11906	GTCGTCGTACACGGGTGGACCGGCGCGGCTCGCGGGCGATTCGCCCGCTGCGCTCGCGCG	11966
OY	611	CGGGCGGCGACCCCTCGCGGCCACCGCACCGGAGCGGGAGCCGCTTGGGCG	656
Db	11966	GCGGGCGCGACACTGGCGGTACCAAGCCGCGCGGCGCGCGGACG	12011

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match	1.48:	Score 61.2:	DB 4:	Length 4411529:
Best Local Similarity	44.48:	Pred. No. 0.0043:		
Matches 287, Conservative	0:	Mismatches 358:	Indels 1:	Gaps 1:

OY	89	ATCCGCGGTTTAAAGAGGGGGCGGGGGCCCTTCCGGGGGCAATCCTTGAGCGACATGGA	148
Db	1619190	ATGTTCCGTAACGGCGGGGGCGGGGGTGTCCGGTTACACCCCGGGCGCCGGAACCGGC	1619131
OY	149	CTACGAGCGGGCTCGGGCGCTGTGGCGGGCCCGAAGACAGAGCCCTGGGGCGGCGGAAGCCA	208
Db	1619130	GCCGGTGGAGCCGGCGGGCGGGTGTGAGAACGGTGGATGAGTAAAGCGCAACGGCGGCC	1619071
OY	209	CTTCCCTCCCCCGGCACCGTAAAGCTCMAAGAACCCGGGGCCCCCGCTGGCGCTCTCCCAAGG	268
Db	1619070	GGCGGCACAAGCGGGTACTGTGGCGCCAGCGCGCCCAAGCGCGGGCGGGCGCTCCCCCGTC	1619011
OY	269	CGGAGAGCCCCCGCGCTTCCCGGGCGGCGCTGGCGGGCAAGAGCCGGGGCTTGTACTCC	328
Db	1619010	CCGCGCAACCCCTGGCGGGTAAAGGGGGTGTGGCGTATGCGGGTGGTGTCTGTGGGATGTC	1618951
OY	329	GGCGGGGGCGCCCCGGGGCAGCAGAGAGAGCTGGGGCGGTTCCGTGCCCTTTCCTG	388
Db	1618950	GGCACCGGCGCGCGGTGGGAGCCGGGGAGACGGCGCGCCGCGGGCGCGGAGAC	1618891
OY	389	TCCGCCCCCGGGCCACCAAGAGCGCGGATTTGGGGGGAGGCTCCCGCAGCGGAGCCGG	448
Db	1618890	CTTAACAGCGGTGCCAAMGGCGCCCGTGGCGGCGACGGCGCAACGGCGCTCGCGGTGG	1618831
OY	449	CTGCAGCCCCCGGCCCAAGTATCAAGGGCGTCTCCCATTTAGACGGGCTTCTGTGGC	508
Db	1618830	GCTGTGGGCCGCTGTTTCGGCAATGGCGGGCGCGCGGCACAMGGGGGACGCGTGGCCAA	1618771
OY	509	GGCGGCCCAAGAGCCTTACGCGCTTGGGCTGGGGGACAAAGGTGGGGCGGCTTCCCGCTGC	568
Db	1618770	GGTGGCGCAGCGCGGACAGCGGCGCTTAGCGCTGGCGCATCGCGCGAGAGGGGACCGGC	1618711
OY	569	CACGCGCTCGAGACCGCGGGAGACCCCAACACATCACTCTGGCGGGCGGCCACCCCTCGC	628
Db	1618710	GGCGCCGGTGTACTGGCGGACACAGTGTGCAATCGCGGCAACAGCGCCCGCGCGCGGC	1618651
OY	629	GCCCAACCGCACCCCGCGGACCTTGGCGGCGACGAGGAGCGAGTGAAGATATAGAGAA	688
Db	1618650	GCCGCGGCGACGGCGGGCGCGGGCGC-ACCGGCGGGCGGACGTCAATGATCGCGGCA	1618592
OY	689	GAGCCCTCTCGGGGGGTGGTGGCGGCTGGGGAGCTCCAGTAAAGGC	734
Db	1618591	AGGGGATACCGCGCGCAACGGCGGCGTGTGGAGCGACGGCGGCC	1618546

```

RESULT 8
US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007_00
: CURRENT APPLICATION NUMBER: US/09/103_840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1

```

```

: SFO ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match	1.3%	Score 58.2	DB 4	Length 4403765
Best Local Similarity	47.7%	Pred. NO. 0.021		
Matches 201; Conservative	0	Mismatches 218	Indels 2	Gaps 1

QY	257	TTCCCTCCAGGGCGGGAGACCCCGCGCTTCCCGGCGCGCTCGGGCGCAAGGGCGGG	316
Db	928834	CGCCGCCAGAGAGCCCGGTCGCCGCTGCTCCGCGCCGCCGCTTGGCGCCGCTGTGCTGG	928893
QY	317	CTTGTACTCCCGCGCGGGGGCGCCGCCGGGAGCAGAGAGAGAGCTGGGGCGTTCGGT	376
Db	928994	CGGCTCCGCCAACCCCCCGAGCGCGCGCATGCGCGCGCGCATCAATCCGAGGGATCCG	928953
QY	377	GCCCTTGGCCCTGTCTCCGCCCCCGGCCACCAAGACCGGCATTGGGGGGGAGCTTCCGC	436
Db	928954	CGCCGCTCCCGCCATGGCGCGCGCTACAGCCGAAAGGGCGCGCGCGGAGACGCGCGCG	929013
QY	437	AGCCGAGACCGGCTGACGCCCCCGGCCCAAGTATCAGCGGCTGCCCCATTGACAGCGG	496
Db	929014	CGCGGGCGCGCGC--CAGCAGGCGGAGAGGCCCAATCGCGCGCGCGCGCATGCGCGCGT	929071
QY	497	CTCTCTCGTGGCGCGGCGCCAAAGAGCCTACGCCCTGGGCTGGGGAGACAGGGTGGCGGC	556
Db	929072	CTCGGTGGCGGCGCGCGCCAGGCCCGCGCGCGCGCGCGAGCCCGCCACCGCCAGAGCCCGAA	929131
QY	557	CTCCCGCGCTGGCACCGGCTGGAGCGCGCGGAGACCCCGACCACTACTCTGCGCGGGGC	616
Db	929132	GCGCGCGGACCAAGCGCCACCGGTCGCCCGGTCGCCCGGCGACCGGTGCGACGCCACC	929191
QY	617	GCCACCCCTCGCGCCACCGCACCGCGCGCGGAGCCTTGGCGCGCGACGAGGAGGAGATGAA	676
Db	929192	AAGCCGCGCGCGCGCGCGGTAGCGCGCGCGCGCAACCGATGAGTGGGGGCAACAGCGAA	929251
QY	677	G 677	
Db	929252	G 929252	

```

RESULT 9
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 2436-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 1.3%; Score 58.2; DB 4; Length 4411529;
Best Local Similarity 47.7%; Pred. No. 0.021;
Matches 201; Conservative 0; Mismatches 218; Indels 2; Gaps 1;

```

```

; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT9p1-F15
;
US-08-232-463-14

Query Match 1.5%; Score 67; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. No. 8.6e-06;
Matches 10; Conservative 233; Mismatches 138; Indels 0; Gaps 0;

QY 784 AGCAGCAGCAGCTGCAGCCAGGAAGAGATCGAGCTGAAGTCAGAGAGACA 843
DB 1432 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1373
QY 844 CGCTCTTCTGCTGATGAACGTATGAAAGCGGATGCACTGTAAGAAAGATACG 903
DB 1372 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1313
QY 904 AGAAGAAAGGCAAGCTGTTTCAGGCGTATGAAGTGAAGAGAGAGAAAGAGAGC 963
DB 1312 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1253
QY 964 TATCTGAGAAATTAAGTGAAGTGCAGCGGAGCTTTCGAGACATCCAGACTCTGC 1023
DB 1252 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1193
QY 1024 CTCCTCAAGCCCTTCTATGTGGCGGAGTGGAAAGGACATAAAGAAATCCCATTTG 1083
DB 1192 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1133
QY 1084 GAAGTCAGAAAGAGACTCCTGTAAGAACTGCTCCTCAATTTCAAAAGTCAAAA 1143
DB 1132 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1073
QY 1144 CAAAACTCTAGACACTCTC 1164
DB 1072 RRRRRATCGCAAGCTCCCTC 1052

RESULT 6
US-09-298-568-3/c
; Sequence 3, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
```

```

; APPLICANT: Bailestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
;
US-09-298-568-3

Query Match 1.4%; Score 62; DB 4; Length 801;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;
Matches 242; Conservative 0; Mismatches 235; Indels 9; Gaps 3;

QY 236 GGAAGCCGGGGCCCCCTGACCTCCCTCCAGAGGCGGAGCCCGCCCTTCCCGGCGG 295
DB 512 GGGACAGGGGGGCCCGCCCGGAGCCCGCGGGAGGGGGCGCCCGCGCGCGG 453
QY 296 CTCGGCGGCAAGGCGCGGGGCTTGTACTCCCGCGGGGCGCCCGCGGAGAGGA 355
DB 452 CTCGGCGCGGGGGCGCGGGAGAAAGCAAGCCCGCGCGCGCGCGGCGGCGG 395
QY 356 AGAGAGCTGGGGGGGTTGGCTTGCCTTGCCTGCGCCCGCGCGGCAAGCAAGCAG 415
DB 394 ---CGCGCGGGCGGATCGCCCGCCCGCTGTTTACCCCGCGGGGGGGCGGCGG 338
QY 416 CATTTGGGGGGGAGCCCTGCGGCGGAGCGGCGGCTCAGCCCGCGGCGCAAGTATAGG 475
DB 337 CGCGGGGGAGCGGGGGCGGAGAGAGGTGCGGGGGCGCGCGCAAAAAAAGTAGGCGT 278
QY 476 GGTGTGCTCCATTCACAGGGGCTCTCTGTCGGGGCGGCGCAAGAGCTACGCTGAGG 535
DB 277 GGGCTAGCCCGTACCCCGCGGGGCTCGGGGCTCGGGGCGCGCGCGCGCG 218
QY 536 TGGGGAAGAGGTGGGGCGGCTTCCCGCTGCGACCG---CTCGAGCCCGGGGAGCC 592
DB 217 TGGCGGCGCGGGGGCGGCTCGCGCTCGCGCGCGGCGGCGGGGGCGCGCGCG 158
QY 593 CCCAGCACTACCTCTCGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
DB 157 CTCGCCGTCGCCCGCGGCGGAGAGCGCGCGGCGGCGGCGGCGGCGGCGG 98
QY 652 TGGCGGCGGAGGCGGAGATGGAAGATATGGAAGAGCCCTCTCGGGGGTGTGGCG 711
DB 97 CGGCTCGCGGGCGGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 38
QY 712 GCTCGG 717
DB 37 GCGCGG 32

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```


QY 236 GGAGCGGGGCCCCCTGCTCTCCAGAGGCGGAGCCCCGCTTCCCGGCGG 295
11 1111 1111 11 1111 1111 1111 1111 1111 1111
Db 512 GGGACAGGGGGCCCCCGGAGCCCCCGGGGAGGGGGCGCGCGCGCGCGG 453
QY 296 CTCGGCGGCAAGGGCGGCTTGTACTCCCGGGCGGCGCGCGCGAGAGA 355
11 1111 1111 11 1111 1111 1111 1111 1111
Db 452 CTCGGCGGCGGCGGCGGAGAGAAAGCAAGGCGCGCGCGCGCGGCG-- 395
QY 356 AGAGACTGGGGGCTTGGCTTGTCCGCTCCCGCGGCGGCGGAGAGAGCGG 415
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 394 ---CCGCGCGGCGGCGATCCCCCGCTTGTACCCCCCGGGGGCGCGG 338
QY 416 CATTGGGGGAGACCTCCGCGAGCGGCTCCAGCCCCCGGCGGCGGAGAGC 475
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 337 CCGCGGAGACCGGGGGCGAGAGAGTTGGGGGCGCGCGGCGGCGGCGG 278
QY 476 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 535
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 277 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 218
QY 536 TGGGAGCAAGGGTGGGGGCGCTCCCGCTGCGGCGGCGGCGGCGGCGG 592
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 217 TGGCGGCGGCGGCGGCGGCGCTCTGCTGCGGCGGCGGCGGCGGCGGCG 158
QY 593 CCGACACTACTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 157 CTCGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98
QY 652 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 711
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 97 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 38
QY 712 GCTCGG 717
11 111
Db 37 GCGCGG 32

RESULT 2
US-08-757-669A-16/c
; Sequence 16, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
US-08-757-669A-16

Query Match 1.6%; Score 70; DB 4; Length 801;
Best Local Similarity 50.8%; Pred. No. 5.9e-07;
Matches 247; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

QY 236 GGAGCGGGGCCCCCTGCTCTCCAGAGGCGGAGCCCCGCTTCCCGGCGG 295
11 1111 1111 11 1111 1111 1111 1111 1111 1111
Db 512 GGGACAGGGGGCCCCCGGAGCCCCCGGGGAGGGGGCGCGCGCGCGG 453
QY 296 CTCGGCGGCAAGGGCGGCTTGTACTCCCGGGCGGCGGCGGCGGAGAGA 355
11 1111 1111 11 1111 1111 1111 1111 1111
Db 452 CTCGGCGGCGGCGGCGGAGAGAGTTGGGGGCGCGCGGCGGCGGCGG 278
QY 356 AGAGACTGGGGGCTTGGCTTGTCCGCTCCCGGCGGCGGCGGCGGCGG 415
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 394 ---CCGCGGCGGCGATCCCCCGCTTGTACCCCCCGGGGGCGCGGCG 338
QY 416 CATTGGGGGAGACCTCCGCGAGCGGCTCCAGCCCCCGGCGGCGGCGGCG 475
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 337 CCGCGGAGACCGGGGGCGGAGAGAGTTGGGGGCGGCGGCGGCGGCGG 278
QY 337 CCGCGGAGACCGGGGGCGGAGAGAGTTGGGGGCGGCGGCGGCGGCGG 218
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 476 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 535
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 277 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 218
QY 536 TGGGAGCAAGGGTGGGGGCGCTCCCGCTGCGGCGGCGGCGGCGGCGG 592
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 217 TGGCGGCGGCGGCGGCGGCGCTCTGCTGCGGCGGCGGCGGCGGCGGCG 158
QY 593 CCGACACTACTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 157 CTCGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98
QY 652 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 711
11 1111 1111 1111 1111 1111 1111 1111 1111
Db 97 GGTGCTGCTTCAACAGGGGCTCTCTGCTGGCGGCGGCGGCGGCGGCG 38
QY 712 GCTCGG 717
11 111
Db 37 GCGCGG 32

RESULT 3
US-09-230-371A-16/c
; Sequence 16, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 182 Seconds
(without alignments)
7366.982 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagtcctgcgaccccgac.....aaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	70	1.6	801	2	US-08-770-379-16
C 2	70	1.6	801	4	US-08-757-669A-16
C 3	70	1.6	801	4	US-09-230-371A-16
C 4	67.2	1.5	936	1	US-08-018-977C-4
C 5	67	1.5	7218	1	US-08-232-463-14
C 6	62	1.4	801	4	US-09-298-568-3
C 7	61.2	1.4	4411529	4	US-09-103-840A-1
C 8	58.2	1.3	4403765	4	US-09-103-840A-2
C 9	58.2	1.3	4411529	4	US-09-103-840A-1
C 10	57.6	1.3	15872	4	US-09-105-537-1
C 11	57.6	1.3	43280	2	US-08-804-227C-1
C 12	56.6	1.3	8438	1	US-07-945-283-1
C 13	56.4	1.3	5852	1	US-07-867-106-2
C 14	56	1.3	4403765	4	US-09-103-840A-2
C 15	55.6	1.3	7898	4	US-08-984-709A-49
C 16	55.2	1.3	5539	4	US-08-628-829-3
C 17	54.6	1.2	7218	4	US-08-232-463-14
C 18	54.4	1.2	13987	2	US-08-804-227C-13
C 19	54.4	1.2	44377	2	US-08-804-227C-7
C 20	54.4	1.2	44377	2	US-08-804-198-1
C 21	54	1.2	1941	5	PCT-US91-07635-5
C 22	52.2	1.2	1325	4	US-09-412-102-3
C 23	52.2	1.2	1325	4	US-09-217-787-3
C 24	52.2	1.2	4257	2	US-08-690-473-1
C 25	52.2	1.2	4257	4	US-09-259-821A-1
C 26	52.2	1.2	4257	4	US-08-843-659-1
C 27	52.2	1.2	12001	1	US-08-458-568A-11

28	51.8	1.2	1227	1	US-08-924-254-1	Sequence 1, Appli
29	51.8	1.2	1227	1	US-08-924-254-3	Sequence 3, Appli
30	51.8	1.2	1227	1	US-09-120-249-1	Sequence 1, Appli
31	51.8	1.2	1227	3	US-09-120-249-3	Sequence 3, Appli
C 32	51.8	1.2	6453	1	US-08-306-691B-14	Sequence 14, Appli
C 33	51.8	1.2	6453	3	US-09-209-668-10	Sequence 10, Appli
C 34	51.8	1.2	6453	4	US-09-356-952-8	Sequence 8, Appli
C 35	51.6	1.2	5253	3	US-09-423-890-7	Sequence 7, Appli
36	50	1.1	3132	2	US-08-224-482-3	Sequence 3, Appli
37	50	1.1	3132	2	US-09-205-921-1	Sequence 1, Appli
38	49.8	1.1	1276	4	US-09-177-325-2	Sequence 2, Appli
39	49.8	1.1	1276	4	US-09-411-812A-2	Sequence 2, Appli
40	49.8	1.1	1276	4	US-09-590-113-2	Sequence 2, Appli
41	49.6	1.1	1225	1	US-08-286-020-1	Sequence 1, Appli
42	49.6	1.1	1225	1	US-08-603-919-1	Sequence 1, Appli
43	49.6	1.1	2186	2	US-08-878-546-9	Sequence 9, Appli
44	49.4	1.1	1241	1	US-07-593-657-6	Sequence 6, Appli
45	49.4	1.1	1241	4	US-08-943-012B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-770-379-16/c
Sequence 16, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770, 379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-770-379-16
Query Match 1.6%; Score 70; DB 2; Length 801;
Best Local Similarity 50.8%; Pred. No. 5.9e-07;
Matches 247; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

```

QY 2752 TTACGAGACCTTGTGCTGGAGATAGACCTTAGGAGGAGGAGGAGAGTGGAA 2811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 ATGGCCAACTTTGTGTGAGTAGAGTTGGCGCGCCAGGGGTTAGGCGATGG 140
QY 2812 ATAGGNG-CTTCCTTTGGCGGCAATGCTACATCTTGAAACAACATGTACCTAAT 2870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 AAGGATGTGTCTCCGCGACAGATGTCTACATCTTGAAACAACATGTACAGAT 80
QY 2871 GACCTTCTCATCTACCTTTGTAATAATA-ATTGTATGTGTACCATTTGGTCC 2923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 GAACCTT---TATTCGCTTAGTGAAGATGATGATGATGATGATGATGATGATGAT 29

RESULT 14
US-09-918-995-15186/c
; Sequence 15186, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15186
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(458)
; OTHER INFORMATION: n.= A,T,C or G
US-09-918-995-15186

Query Match 1.5%; Score 66.8; DB 9; Length 458;
Best Local Similarity 91.0%; Pred. No. 3.2e-07;
Matches 71; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1488 TCAGCTTGAGCTTCTCTCTTTGGAGGAGCAGCTAGTAGAGCTTAAAGGACCAAT 1547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 TCTTTTTCACAGTCTCTCTTTGGAGGAGCAGCTAGTAGAGCTTAAAGGACCAAT 312
QY 1548 CCTTCAGACCTTTGGAG 1565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 CCTTCAGACCTTTGGAG 294

RESULT 15
US-09-976-740-48
; Sequence 48, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547

```

```

; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Oryzotia cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1895)
US-09-976-740-48

Query Match 1.5%; Score 64.8; DB 9; Length 2561;
Best Local Similarity 49.4%; Pred. No. 4.3e-06;
Matches 252; Conservative 0; Mismatches 252; Indels 6; Gaps 3;

QY 229 AGCTCAAGAGACCGGGGCCCCCTGCTCTCTCCAGAGGCGGAGCCCGCCCTTCCC 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 ACCGCAACGGGGCGGCTGCACGCCCGCGGAGGACACCCCGCGCGCGCGCGCG 585
QY 289 CGCGCGGCTGCGGCGGCAAGGCGCGGCGCTTTACTCCGCGGCGGCGCGCGCGCG 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
QY 349 AGCAGAGAGAGACTGCGGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705
QY 409 AAGCGGCAATTTGGGGGGAGCTGCGGCAAGCGGAGCGGCGCT---CAGTCCCGGCGCA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 CCGCGCGCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765
QY 466 AGTATCAGGCGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 AGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 825
QY 526 CGCGCTGCGGCGGCGGCAAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
QY 586 CGGAGACCCGACACACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 CGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 944
QY 646 GAGCGCTGCGGCGG---CAGCGAGGCGAGATGAGAGATGAGAGAGAGCGCTTCCGAGG 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 CCAACCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 704 TGGTGGCGGCTGCGGAGCGCTTCCAGTCAAGG 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1034

```

Search completed: July 14, 2003, 20:15:40
Job time : 672 secs

Db 245 GCCAGATTTCTTGTCTCCATCATTTTAATATGCAAACTGTGGCTTTCCTACTCT 304
OY 3641 TACCATGTGATTTCTTGTACGTGTCCAAAAGAAAA 3682
Db 305 GACCATGTATGTCTTAAATATGCCAAAAGAAAA 346

RESULT 12
US-10-106-698-3799

Sequence 3799, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
PRIOR FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 3799
LENGTH: 196
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (7)..(8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (39)..(39)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (70)..(70)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (76)..(76)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (78)..(78)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (128)..(128)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (143)..(143)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (146)..(146)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (180)..(180)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc-feature
LOCATION: (196)..(196)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-3799

Query Match 4.0%; Score 176.2; DB 9; Length 196;
Best Local Similarity 93.2%; Pred. No. 6.8e-37;
Matches 178; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3810 GGGAGCTCCCGAGTTTCTGACTGAGTAGAGTCCAGAGGCTGCTATCTGG 3869
Db 5 GGNNTTCCCGAGTTTCTGACTGAGTAGAGTCCAGAGGCTGCTATCTGG 64
OY 3870 CCAAGTTAAGTATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3929
Db 65 CCAAGTTAAGTATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124

OY 3930 TTGCCCCCTCTTCTAGGTTAAATTCCTTGTGACTTGTGAGAGAGGCTTGAC 3989
Db 125 TTGCCCCCTCTTCTAGGTTAAATTCCTTGTGACTTGTGAGAGAGGCTTGAC 184
OY 3990 AGTAGATTAGC 4000
Db 185 AGTAGATTAGC 195

RESULT 13
US-09-917-800A-737/c

Sequence 737, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendlick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 737
LENGTH: 565
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1102820
US-09-917-800A-737

Query Match 2.5%; Score 107.2; DB 10; Length 565;
Best Local Similarity 62.6%; Pred. No. 5.3e-18;
Matches 259; Conservative 0; Mismatches 133; Indels 22; Gaps 5;

OY 2512 CTGTGATACAGTCATGTGGGAAGGATTTGGCTGTGATTTTTCACATTAATGAT 2571
Db 422 CTGGACAGTGTGATGTGGGAAGACATGTGGCTG-CTTACCTTCTTAATTATGGGT 364
OY 2572 AACAAATTTCTTACGCTCAAAAACCAAAATCTTGAAGAAAGGAGGTGATAGT 2631
Db 363 AACGTAAGTCTTCCACACCCCAACAGAGAAATGTTAGAAAAGGGTTGGGCTTAG 304
OY 2632 TTCAGAACACCTTACAGCTGTAAACAAAGCACTTAGTATTGGAGTGCACCAAAAC 2691
Db 303 TTTCTG-----AAGGATACAGTTGTCTACTTAGGAGTGCACACACAGGT 260
OY 2692 CTGTATAATGTCTTGTATGATACATCTTTCAGATATTCTTCATTTGGGCTGATCCT 2751
Db 259 CTGTGAGATGTCTGTGTCATGTACATCTTCAAGTAAGTATCTTGTGGGCTTCAGCCT 200

QY 3867 TGG 3869
Db 539 TTG 541

RESULT 7

US-09-918-995-34149
; Sequence 34149, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 34149
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34149

Query Match 8.9%; Score 389.6; DB 9; Length 411;
Best Local Similarity 98.8%; Pred. No. 6.4e-94;
Matches 403; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 2179 GGCAGATTGAGAACTTTGAGGGGCTCTTTCTATACATTTGCTATGTAAAGGGGTAAA 2238
Db 1 GGCAGATTGAGAACTTTGAGGGGCTCTTTCTATACATTTGCTATGTAAAGGGGTAAA 60
QY 2239 AGGGCTCTCTCATATACATGTGGAAGATGAAACACCCCTTCTTAAAGCTGTGCT 2238
Db 61 AGGGCTCTCTCATATACATGTGGAAGATGAAACACCCCTTCTTAAAGCTGTGCT 120
QY 2299 GCATGACACTCTTCTCAACCCGTGACACCTCTTATAGTGGATATGATTTTAAAC 2358
Db 121 GCATGACACTCTTCTCAACCCGTGACACCTCTTATAGTGGATATGATTTTAAAC 180
QY 2359 CTAATAATTAACAACAACCTCACCATG-AACCTTAGGACAGAAAGGAATGCAAGTG 2417
Db 181 CTAATAATTAACAACAACCTCACCATGAGCTTAGGACAGAAAGGAATGCAAGTG 240
QY 2418 AAGGATGAAGCAAGCATCTTCACAGTAGAAGAAACATCGAGAGTTGGTAGATAC 2477
Db 241 AAGGATGAAGCAAGCATCTTCACAGTAGAAGAAACATCGAGAGTTGGTAGATAC 300
QY 2478 TGTCTGAAAAGATGATTCTCATTTGAAACTATTCTGTATACAGTCATGTGGGAAGGA 2537
Db 301 TGTCTGAAAAGATGATTCTCATTTGAAACTATTCTGTATACAGTCATGTGGGAAGGA 360
QY 2538 TGTTCGCTGTGATATTTTTCAGTTAATGATGAACAATTTCTTAC 2585
Db 361 TGTTCGCTGTGATATTTTTCAGTTAATGATGAACAACAATTTCTTAC 408

RESULT 8

US-09-918-995-30064
; Sequence 30064, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 30064
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30064

Query Match 8.2%; Score 357; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 3.3e-85;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 919 AGCTGTTTCAGGCTATGAACTGAGAGACAGAGAAACAGACTATCTGAGAAATTA 978
Db 9 AGCTGTTTCAGGCTATGAACTGAGAGACAGAGAAACAGACTATCTGAGAAATTA 68
QY 979 AACTGAGATGCGACCCGAGCTTCCGAGACATCCGACCTCCCTCCCAAGCCCTCT 1038
Db 69 AACTGAGATGCGACCCGAGCTTCCGAGACATCCGACCTCCCTCCCAAGCCCTCT 128
QY 1039 CATGTGGCGGAGTGAAGAGACATTAAGAAATCCCATTTGAAGTACAGAAAGAA 1098
Db 129 CATGTGGCGGAGTGAAGAGACATTAAGAAATCCCATTTGAAGTACAGAAAGAA 188
QY 1099 AGACTCTGTTAAAAAGCTGCTCTGAATTTCAAAAGTCAAAACAAACCTCTAAC 1158
Db 189 AGACTCTGTTAAAAAGCTGCTCTGAATTTCAAAAGTCAAAACAAACCTCTAAC 248
QY 1159 ACTCTCTATTAAGAGAACCCCTGTGTCTTATCTGAACCTGTTTGAAGTGAAT 1218
Db 249 ACTCTCTATTAAGAGAACCCCTGTGTCTTATCTGAACCTGTTTGAAGTGAAT 308
QY 1219 TGAGAGCCCAAGAACCCCAAGAACCCCGTCTTCAAGTGAACCCCAAGAAC 1275
Db 309 TGAGAGCCCAAGAACCCCAAGAACCCCGTCTTCAAGTGAACCCCAAGAAC 365

RESULT 9

US-09-917-800A-16/c
; Sequence 16, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Menodick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 590

OY	857	GATTGAACGTATGTGAAAGCGGATGACCTGGTAAAGAAAGTATACGAAAGAAAGGCA	916
Db	242	GATTGACGTATGTGAAAGCGGATGACCTGGTAAAGAAAGTATACGAAAGAAAGGCA	301
OY	917	CAAGCTGTTTCAGGGCTATGTGAACGTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAAT	976
Db	302	CAAGCTGTTTCAGGGCTATGTGAACGTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAAT	361
OY	977	TAAACTGGAGTCCACGCGGAGCTTTCGAGACATCCAGACTCTGCCTCCCAAGCCTT	1036
Db	362	TAAACTGGAGTCCACGCGGAGCTTTCGAGACATCCAGACTCTGCCTCCCAAGCCTT	421
OY	1037	CTCATGTGGGGGAGTGGAAAGGAGCATAAAGAAATCCCATTTGGATGTGACAGAAAG	1096
Db	422	CTCATGTGGGGGAGTGGAAAGGAGCATAAAGAAATCCCATTTGGAGTACGAAAG	481
OY	1097	AAGACATCTCTTAAAAAGCTGCGCTCGATTTTCAAAAGTCAAAACAAAACCTCTAA	1156
Db	482	AAGACATCTCTTAAAAAGCTGCGCTCGATTTTCAAAAGTCAAAACAAAACCTCTAA	541
OY	1157	GCACCTCTCTTAAAGAGAAACCTGTGTTCTTATCTGAACCTGTTTGTAAACGTGA	1216
Db	542	GCACCTCTCTTAAAGAGAAACCTGTGTTCTTATCTGAACCTGTTTGTAAACGTGA	601
OY	1217	ATTGAGGAGGCCAAGAAACCCCAAGAAACCCCGGCTCTCACTGGACACCCCAAGACT	1276
Db	602	ATTGAGGAGGCCAAGAAACCCCAAGAAACCCCGGCTCTCACTGGACACCCCAAGACT	661
OY	1277	CTCCACTCTCCCAAAAGGGAGCCAGACCCATCCCAAGAGAGAAAGCCTTCTCAAGTGAAT	1336
Db	662	CTCCACTCTCCCAAAAGGGAGCCAGACCCATCCCAAGAGAGAAAGCCTTCTCAAGTGAAT	721
OY	1337	AGAAAGATTTCGCCGTACCTTTCACACAGAAATGTAATTTGTGTCTTGGACACGACTCC	1396
Db	722	AGAAAGATTTCGCCGTACCTTTCACACAGAAATGTAATTTGTGTCTTGGACACGACTCC	781
OY	1397	CCCAATCACCGTTACCATTAAGGGAATCTCTCCAAAGAAAGAGAGCATGTAGCAAG	1453
Db	782	CCCAATCACCGTTACCATTAAGGGAATCTCTCTCCAAAGAAAGAGAGCATGTAGCAAG	838

RESULT 2 917-800A-725
US-09-917-800A-725
Sequence 725, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendlick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michel
APPLICANT: Gene Logic, Inc.
TITLE OR INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917.800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459

```

: PRIOR FILING DATE: 2001-07-09
:
: NUMBER OF SEQ ID NOS: 1740
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 725
: LENGTH: 615
: TYPE: DNA
: ORGANISM: Rattus norvegicus
:
: FEATURE:
:
: OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1102093
: NAME/KEY: misc.feature
: LOCATION: (1)..(615)
: OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-725

```

Query Match	12.6%	Score 551.6	DB 10	Length 615
Local Similarity	94.6%	Pred. No. 2.9e-137		
Matches 582	Conservative 0	Mismatches 30	Indels 3	Gaps 1
QY	112	CGGCCCTGCGGGGCAATCTGAGCAGCACTGGACTAGAGCGGGCTGGCGCTGG	171	
Db	1	CGGCCCTGCGGGGCAACCCCGAGCAGCGACTGACTACAGAGGGGCTGGCGCTGG	60	
QY	172	GGGGGCGCGAAGAGAGGCTGGGGGGGCGAAAGCCACTTCTCCCGCGACCGCTAAGC	231	
Db	61	GGGGGCGCGAAGAGAGTCCGGGGGGGCGAAGCCACTTCTCCCGCGACTGTGAAGC	120	
QY	232	TCAAGGAGCCCGGGGCCCCCGCTGGGCTCTCTCCAGGGGCGAGCCCGCCTTCCCGG	291	
Db	121	TCAAGGAGCCCGGGGCCCCCGCTGGGCTCTACCCAGGGGCGAGCCCGCGCCTTCCAG	180	
QY	292	CCGGCT---GGGGGCGAAGGGGCGGGGCTGTACTCCGGGCGGGGCGGGCCCCGGGC	348	
Db	181	CTGGCTGGGGGCGGCGAAGGCGGGGTTTGTACTCCCGGCGGGGCGGGCCCCGGGC	240	
QY	349	AGCAGGAAGAGAGCTGGGGGCTTCCGTGCTGCTTCCGCTTCCGCGCCCGGACCAAGC	408	
Db	241	AGCAGGAAGAGAGTGGGGGCTTCCGTGCTGCTTCCGCTTCCGCGCCCGGACTCAAAAC	300	
QY	409	AAGCGGCAATTGGGGGGGAGCCTGCGCAGCCGGAGCCGCTGCAGCCCCGGGCCAAGT	468	
Db	301	AAGCGGCAATTGGGGGGGAGCAGTCGAGCCGGGCTGGGCTGCAGCCCCGGGCCAAGT	360	
QY	469	ATCAGGCGGCTGCTGCCATTCAGAGAGGGCTCTGCTGGGGGCGGACCAAGAGCTTAGC	528	
Db	361	ATCAGGCGGCTGCTGCCATTCAGAGAGGGCTCTGCTGGGGGCGGACCAAGAGCTTAGC	420	
QY	529	CCTGGGCTGGGGACAAAGGTGGGGCGGCTCCCGCGTGCACCGCTCGAGCCCGGCGG	588	
Db	421	CCTGGGCTGGGGACAAAGGTGGGGCGGCTCCCGCGTGCACCGCTCGAGCCCGGCGG	480	
QY	589	GAGCCCGACACTACCTCTGCGCGGGGCGGACCCCTGGCGCGACCGCACCGGCGGGGA	648	
Db	481	GAGCCCGACACTACTCTTGCCTGGGGCGGACCCCTGGCGCGACCGCACCGGCGGGGA	540	
QY	649	CCCTGGCGGCGAGCGAGGAGATGGAAGATATGAGGAAGAGCCCTCTCGGGGGGTGGT	708	
Db	541	CCCTGGCGGCGAGTGGAGGAGATGGAAGATATAAGGAAGAGCCCTCTCGGGGGGTGGG	600	
QY	709	GGGGCTGGGGAGCCT 723		
Db	601	GCNGCTGGGGAGCCT 615		

RESULT 3
US-09-736-457-1554
: Sequence 1554, Application US/09736457
: Patent NO. US20020168637A1
: GENERAL INFORMATION
: APPLICANT: Wang, Tongtong
: APPLICANT: Banquet, Chantanya S.
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Vedvick, Tom

Db 346 -HOKOSRIPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENIDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRLQ 536
Db 402 --TELKALRLNFOROLRO-EVYVCMRDTALETALNAKAYKRSKQSLREARITTEK 458
QY 537 RM---YKKKGIQSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQEYLSIIQIARD 485

RESULT 13

US-09-535-008-69
; Sequence 69, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1678
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-69

Query Match 5.3%; Score 171; DB 4; Length 1678;

Best Local Similarity 19.9%; Pred. No. 0.0023;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKEBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARQOEESWGSVPLPCPPA- 108
Db 1 MSTPDDPL-----GSTRPGRSPGPGSPGAMLGPSGSPGSAHSMWG---PSGPPSAG 53
QY 109 --TKQAGIGEP-----AAAGACSPRPXYQAV--LPITGSLVAAKEPTPW 152
Db 54 HPITGPGGYPQDNNHQMHRPMESMHEKMSDDPRNQMGKMGSRGAGMGPPSPM 113
QY 153 AGDKGAASPAATASDPAGRPPLPGLPPLATATAGTAAEGMRKSRKSPGLGGGG 212
Db 114 DQHSQGYPSPLGSEHASSVPASGSSG-----QMSGP--GAPL 154
QY 213 SGASSQAACIKQILLDLIEQOOOOLAKEREIEELKSE--RDTLLARIERMERMOL 270
Db 155 DGADPOA-----LGOONRGPTPENQMLHQLRAQIMAYKMLARGPLPHLQW 202
QY 271 VKKDNEKERHKLQGYETEREBTELSEKIKLECEBELSETSOTLPKPPSCGRSGK-- 328
Db 203 A-----VOGKRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPGSTERKTYPVKKLAPESKYVTKPKHSPIKEEPCGSLSET 375
Db 236 PGCGPGRGAPRMYSRPHGMGPNMPPRGSSGVPFGMPGQ---PRGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPRKPRSSVDTPRLSTPOKGPSTHREKAFSSIEDLPYSTTEMYLGR 435
Db 293 AAPTSTPQKILIPQIPGRSPAPPAVPAPMPPOTGSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVAVCL-----MPSVAGETSVLAVPSW 478
Db 346 -HOKOSRIPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENIDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRLQ 536
Db 402 --TELKALRLNFOROLRO-EVYVCMRDTALETALNAKAYKRSKQSLREARITTEK 458

QY 537 RM---YKKKGIQSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQEYLSIIQIARD 485

RESULT 14

US-09-535-008-65
; Sequence 65, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 1679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-65

Query Match 5.3%; Score 171; DB 4; Length 1679;

Best Local Similarity 19.9%; Pred. No. 0.0023;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKEBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARQOEESWGSVPLPCPPA- 108
Db 1 MSTPDDPL-----GSTRPGRSPGPGSPGAMLGPSGSPGSAHSMWG---PSGPPSAG 53
QY 109 --TKQAGIGEP-----AAAGACSPRPXYQAV--LPITGSLVAAKEPTPW 152
Db 54 HPITGPGGYPQDNNHQMHRPMESMHEKMSDDPRNQMGKMGSRGAGMGPPSPM 113
QY 153 AGDKGAASPAATASDPAGRPPLPGLPPLATATAGTAAEGMRKSRKSPGLGGGG 212
Db 114 DQHSQGYPSPLGSEHASSVPASGSSG-----QMSGP--GAPL 154
QY 213 SGASSQAACIKQILLDLIEQOOOOLAKEREIEELKSE--RDTLLARIERMERMOL 270
Db 155 DGADPOA-----LGOONRGPTPENQMLHQLRAQIMAYKMLARGPLPHLQW 202
QY 271 VKKDNEKERHKLQGYETEREBTELSEKIKLECEBELSETSOTLPKPPSCGRSGK-- 328
Db 203 A-----VOGKRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPGSTERKTYPVKKLAPESKYVTKPKHSPIKEEPCGSLSET 375
Db 236 PGCGPGRGAPRMYSRPHGMGPNMPPRGSSGVPFGMPGQ---PRGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPRKPRSSVDTPRLSTPOKGPSTHREKAFSSIEDLPYSTTEMYLGR 435
Db 293 AAPTSTPQKILIPQIPGRSPAPPAVPAPMPPOTGSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVAVCL-----MPSVAGETSVLAVPSW 478
Db 346 -HOKOSRIPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENIDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRLQ 536
Db 402 --TELKALRLNFOROLRO-EVYVCMRDTALETALNAKAYKRSKQSLREARITTEK 458
QY 537 RM---YKKKGIQSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQEYLSIIQIARD 485

```

OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTPKHSPIKEPCGSLSET 375
DB 236 PGPGRGPAPARNYSRPHGMGMPMPGPGSVPPGPGQ---PPGPPKPMPEGGMANA 292
OY 376 VCKRELISOETPERKPRSSVDTPRLSTPOKGPSTHPEKAFSSIEDLYLSTTEMYLCR 435
DB 293 AAPSTPOKILPPQPTGRPSAPPAVPAPASVMPPOGOSPOQAPAPMVL----- 345
OY 436 WHOPPSPLP-----LRSSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
DB 346 -HOKOSRITPIOKPRGLDPEVILQERREYRLQARIARHIOELENLPGSLADLRTKAT--- 401
OY 479 RDSVPELRDPNPSDLENDLSVFSKRHAKE--LDEKRRKMDIQIRIORILOL 536
DB 402 --TELKALRLNFQOLRO-EVYVCMRDTALETNALNAKAYKRSKROSLREARITKLEK 458
OY 537 RM---YKKGIOESEPEVTSEFPEDD 560
DB 459 OOKIEORRRKROKHQEYLNSTILOHAKD 485

```

```

RESULT 11
US-09-535-008-75
; Sequence 75, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGL1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 75
; LENGTH: 1649
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-535-008-75

```

```

Query Match
Best Local Similarity 5.3%; Score 171; DB 4; Length 1649;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

OY 52 LKEPPPLASSQGSPPAPSPAGCGKRGILL--PAGAARQOEESMGSGVPLPCPPA- 108
DB 1 MSTDPPL-----GGTPRPGSPGPGSPGAMLGSPSPGSAHSMKG--PSPGPPSAG 53
OY 109 --TKOAGIGEP-----AAAGACSPRPKYAV--LP1QTSLVAAAKETPM 152
DB 54 HPIPTQGRGGYPODNHOMHKRMESMHEKMSDDPRYNOMKGMNGSGHAGMGPSPM 113
OY 153 AGDGGAAPATAADAPAPPLPLPDPPLAATAGTAAASEGRKMSKSPSLGGGG 212
DB 114 DQHSQGPSPGLGSEHASSVPASGSSGP-----QMSGGP--GGAPL 154
OY 213 SGASSQAACIKOTLLDLLEQOOQOOLAKKEIEELKSE--RDTLLARIERMERMOL 270
DB 155 DGADPQA-----LGOQNRGPTFPNONOLHOLIAQIMAYKMLARQPLDHIOM 202
OY 271 VKKDNEKERHKLFGYETEERETELSEKIKLEQPELSESTQTLPPKPSGSGSGK- 328
DB 203 A-----VOGRKP-----MPGMOOQMPPLPPPSVATGPGPG 235
OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTPKHSPIKEPCGSLSET 375
DB 236 PGPGRGPAPARNYSRPHGMGMPMPGPGSVPPGPGQ---PPGPPKPMPEGGMANA 292
OY 376 VCKRELISOETPERKPRSSVDTPRLSTPOKGPSTHPEKAFSSIEDLYLSTTEMYLCR 435

```

```

DB 293 AAPSTPOKILPPQPTGRPSAPPAVPAPASVMPPOGOSPOQAPAPMVL----- 345
OY 436 WHOPPSPLP-----LRSSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
DB 346 -HOKOSRITPIOKPRGLDPEVILQERREYRLQARIARHIOELENLPGSLADLRTKAT--- 401
OY 479 RDSVPELRDPNPSDLENDLSVFSKRHAKE--LDEKRRKMDIQIRIORILOL 536
DB 402 --TELKALRLNFQOLRO-EVYVCMRDTALETNALNAKAYKRSKROSLREARITKLEK 458
OY 537 RM---YKKGIOESEPEVTSEFPEDD 560
DB 459 OOKIEORRRKROKHQEYLNSTILOHAKD 485

```

```

RESULT 12
US-09-535-008-71
; Sequence 71, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGL1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 71
; LENGTH: 1650
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-535-008-71

```

```

Query Match
Best Local Similarity 5.3%; Score 171; DB 4; Length 1650;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

OY 52 LKEPPPLASSQGSPPAPSPAGCGKRGILL--PAGAARQOEESMGSGVPLPCPPA- 108
DB 1 MSTDPPL-----GGTPRPGSPGPGSPGAMLGSPSPGSAHSMKG--PSPGPPSAG 53
OY 109 --TKOAGIGEP-----AAAGACSPRPKYAV--LP1QTSLVAAAKETPM 152
DB 54 HPIPTQGRGGYPODNHOMHKRMESMHEKMSDDPRYNOMKGMNGSGHAGMGPSPM 113
OY 153 AGDGGAAPATAADAPAPPLPLPDPPLAATAGTAAASEGRKMSKSPSLGGGG 212
DB 114 DQHSQGPSPGLGSEHASSVPASGSSGP-----QMSGGP--GGAPL 154
OY 213 SGASSQAACIKOTLLDLLEQOOQOOLAKKEIEELKSE--RDTLLARIERMERMOL 270
DB 155 DGADPQA-----LGOQNRGPTFPNONOLHOLIAQIMAYKMLARQPLDHIOM 202
OY 271 VKKDNEKERHKLFGYETEERETELSEKIKLEQPELSESTQTLPPKPSGSGSGK- 328
DB 203 A-----VOGRKP-----MPGMOOQMPPLPPPSVATGPGPG 235
OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTPKHSPIKEPCGSLSET 375
DB 236 PGPGRGPAPARNYSRPHGMGMPMPGPGSVPPGPGQ---PPGPPKPMPEGGMANA 292
OY 376 VCKRELISOETPERKPRSSVDTPRLSTPOKGPSTHPEKAFSSIEDLYLSTTEMYLCR 435
DB 293 AAPSTPOKILPPQPTGRPSAPPAVPAPASVMPPOGOSPOQAPAPMVL----- 345
OY 436 WHOPPSPLP-----LRSSPKKEETVARCL-----MPSVAGETSVLAVPSW 478

```

Db 114 DOHSGCYPSPLGSGEHSASPVPASGSPG-----OMSSGP--GCAPL 154
QY 213 SGASSOACLKQIILLQDLIEOQOOLAKEKEIEIKSE--RDTLLARIEMERMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNOMQLHQAQIMAYMLRAGOPLDPHLOM 202
QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGK-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPFGSTERKTPVKKLAPESKYKTTPKSPKIEEPCGSLSET 375
Db 236 PGPGPGPAPRYNSPHCMGMPMPGPGSVPPMPGQ---PCGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPEKPSVSDTPPRLSTPOKGSTHKEKAFSEIEDLYLSTTEMYLOR 435
Db 293 AAPTSTPQKLIPOQIPGRSPAPPAVPAPSPVMPQOTSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVARCL-----MPSSVAGETSVLAVPSW 478
Db 346 -HOKOSRITPIQKPRGLDPEVILQEREYRLQARIARIQLELNPGLSLAGDLTKAT--- 401
QY 479 RDSVPLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQIRIQRILORLOL 536
Db 402 --TELKALRLNFQRLRO--EYVVCMRBDTALETALNAKAYKRSKROSILREARITEKLEK 458
QY 537 RM---YKKKIOESEPEVTSFPEPDD 560
Db 459 QOKIEQERRRROKHQEYLSILOHARD 485

RESULT 9

US-09-535-008-67
; Sequence 67, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 1646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-67

Query Match 5.3%; Score 171; DB 4; Length 1646;

Best Local Similarity 19.9%; Pred. No. 0.0022;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKRPBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARPGQOEESMGSVPLPCPPPA- 108
Db 1 MSTPDDPPL---GGRTRPGSPGPGSPGAMLPSPGSPGSAHSMWG---PSPGPPSAG 53
QY 109 --TKAGIGGEP-----AAAGAGCSRPXYQAV--LPITGSLVAANAAPETPM 152
Db 54 HPIPTGPGGYPDNDNHNHKKPMESNHEKMSDPRYNQMKGMGMSGGHAGGPPSPM 113
QY 153 AGDGGAAPPAATASDPAGRPPLPLPGPPLAPATAGTAASEGKWSMKRSPLGGGG 212
Db 114 DOHSGCYPSPLGSGEHSASPVPASGSPG-----OMSSGP--GCAPL 154
QY 213 SGASSOACLKQIILLQDLIEOQOOLAKEKEIEIKSE--RDTLLARIEMERMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNOMQLHQAQIMAYMLRAGOPLDPHLOM 202

QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGK-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPFGSTERKTPVKKLAPESKYKTTPKSPKIEEPCGSLSET 375
Db 236 PGPGPGPAPRYNSPHCMGMPMPGPGSVPPMPGQ---PCGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPEKPSVSDTPPRLSTPOKGSTHKEKAFSEIEDLYLSTTEMYLOR 435
Db 293 AAPTSTPQKLIPOQIPGRSPAPPAVPAPSPVMPQOTSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVARCL-----MPSSVAGETSVLAVPSW 478
Db 346 -HOKOSRITPIQKPRGLDPEVILQEREYRLQARIARIQLELNPGLSLAGDLTKAT--- 401
QY 479 RDSVPLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQIRIQRILORLOL 536
Db 402 --TELKALRLNFQRLRO--EYVVCMRBDTALETALNAKAYKRSKROSILREARITEKLEK 458
QY 537 RM---YKKKIOESEPEVTSFPEPDD 560
Db 459 QOKIEQERRRROKHQEYLSILOHARD 485

RESULT 10

US-09-535-008-2
; Sequence 2, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-2

Query Match 5.3%; Score 171; DB 4; Length 1647;

Best Local Similarity 19.9%; Pred. No. 0.0022;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKRPBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARPGQOEESMGSVPLPCPPPA- 108
Db 1 MSTPDDPPL---GGRTRPGSPGPGSPGAMLPSPGSPGSAHSMWG---PSPGPPSAG 53
QY 109 --TKAGIGGEP-----AAAGAGCSRPXYQAV--LPITGSLVAANAAPETPM 152
Db 54 HPIPTGPGGYPDNDNHNHKKPMESNHEKMSDPRYNQMKGMGMSGGHAGGPPSPM 113
QY 153 AGDGGAAPPAATASDPAGRPPLPLPGPPLAPATAGTAASEGKWSMKRSPLGGGG 212
Db 114 DOHSGCYPSPLGSGEHSASPVPASGSPG-----OMSSGP--GCAPL 154
QY 213 SGASSOACLKQIILLQDLIEOQOOLAKEKEIEIKSE--RDTLLARIEMERMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNOMQLHQAQIMAYMLRAGOPLDPHLOM 202
QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGK-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVATGPGPG 235

```

QY 159 AASPAATASDPAGPPPLPGLPPPLAPATAGTLAASEGHW-----KSMKSPILGGGG-- 213
DB 788 AEPRLATEAGTGPRLPLPSVY--SPSOGAGPLPSSEASAPADPALDPSPTPATGGEV 845
QY 214 GASSQAACLOKILLDLEOQQOOLQAKEIEE-----LKSRDPL 257
DB 846 SAILASALNG-----SSSPFEVAPSSSEDEDTAETSGIFDTSSDGLQANRPV 896
QY 258 LARLERMERRM-----OLVKRD-----NEKEHKLPGYTEIEREE 293
DB 897 VPAPRSLOKQVGPDPDLSDIDIPSSADGVEVFPSSATGSGGQPRALDSGYDENYES 956
QY 294 TELSEKIKLE-CQPE-----LSETSOTLPPKPPSCGSGKHRR-----KSPGSTERKT- 342
DB 957 PEVYLKAOGCEPOAFELASEGEGPPETRLSTLSGLNKNRPYRDATYSDELAENE 1016
QY 343 ----PVKKL-----APFSKYTKTPKHSPIKEEPCSLSETVCRELRSOETPEKPRS 392
DB 1017 ATSGPBEKCGGDAPPELGLPSTGCP-----SBOYCLRPQVSGEAGSGSGPG 1063
QY 393 SVTPPPRL-----STPKGPST-----HPKKAFTSEIEDLPYLSTTEMYLCRMHOPRP 441
DB 1064 EV-LPRLDLESSPE--PSTCPGLVPEPPPOGPAKYRPPSPSCSOFLL----- 1113
QY 442 SPLPLRESSPKKEETVACRLMSSVAGETSVLAAPSWDH-----SVEPLR----- 487
DB 1114 TPVPLRSEGNSE-----FOGPPGLSGPAPQRMGPGTPAPRLRLPLPA 1162
QY 488 --DPNSSLLENLDVSFSKRAKLELDEKRRKMDIR-----IREORT 530
DB 1163 ALEGPEEEDSESD-----ESDELR-CYVPEPSEDEEAPAVPVVAESQS 1213
QY 531 LORLOLMYKKKGIOS-----EPEVTSFPEPDEDESIMITPLPVVAFGRLP--K 581
DB 1214 ABNLMSLLKMBLSLSETECEDLERKKAVSFF--DVTYVLFQESPTRELGPFFGAK 1270
QY 582 LTPONF 587
DB 1271 ESPPTF 1276

RESULT 7
US-09-535-008-63
; Sequence 63, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-63

Query Match
Best Local Similarity 19.3%; Score 171; DB 4; Length 543;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

QY 52 LKEGPPLASSOGSPAPSPAGCGGKGLLL--PAGAAPGOEESWGSGVPLPCPPPA- 108
DB 1 MSTPDPPL-----GCTPRPGSPGPGSPGAMLGSPSPSANSMMG---PSPGPPSAG 53
QY 109 --TKAGIGGP-----AAAGAGCSPPRKYQAV--LPITQGSILVAAAKEPTPW 152

```

```

DB 54 HPITPQGGGATPQDNNHQMHRPMESNHEKMSDDPRYNQMGKMRSGHAGCPPPSPW 113
QY 153 AGDKGGAAPATASDPAGPPPLPGLPPPLAPATAGTLAASBGRKMSRKSPILGGGG 212
DB 114 DOHSQGYSPPLSGSEHNASPVPAAGPSSGP-----QMSGP-GGAPL 154
QY 213 SGASSQAACLOKILLDLEOQQOOLQAKEIEELASE--RDTLLARIEREMRMOL 270
DB 155 DGADPQA-----LGQONRGPTPNQOLQRLQAINAYKMLARGOPLPDHLM 202
QY 271 VKKONERKHLFOGYTEREETELSEKIKLECOPELSTOTLPPKPPSCGSGKG-- 328
DB 203 A-----VOGKRP-----MGMOQAMTLLPSPVSATGREGPG 235
QY 329 -----HKRSPPGSTERTKTPVKKLAPESKYTKTPKHSPIKEEPCSLSET 375
DB 236 PGPGPGPAPPNYSRPHGGMNMPGPGSVPPGMPQ---PGGPPKPMPEGPMA 292
QY 376 VCKRELRSOETPEKPPSSVDTPPRLSTPKGPSTHPEKAFSSIEDLPYLSTTEMYLCR 435
DB 293 AAPTSTPQKILPPQPTGRSPAPPAVPAPASVMPQOTSPQAPAPMVP----- 345
QY 436 WHPPSPPL-----LRESSPKKEETVACNCL-----MPSVAGETSVLAAPSW 478
DB 346 -HOKOSRITPQKRGDLDEVEILOEREYRLQARIHQLELNLGSLAGLRTKAT--- 401
QY 479 RDSVPLDNPNSDLENDVSFSKRAKLE--LDEKRRKMDIRQIRQRILOQL 536
DB 402 --TELKALNLPNQRLQ--EYVVCMKRDTALETALNAKVRKRSKRSRIAREARTKLEK 458
QY 537 RM---YKKKGIOSPEVTSFPEPDD 560
DB 459 QOKIEQRKRQKHQEYLSILOHAKD 485

RESULT 8
US-09-535-008-61
; Sequence 61, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-61

Query Match
Best Local Similarity 19.9%; Score 171; DB 4; Length 577;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

QY 52 LKEGPPLASSOGSPAPSPAGCGGKGLLL--PAGAAPGOEESWGSGVPLPCPPPA- 108
DB 1 MSTPDPPL-----GCTPRPGSPGPGSPGAMLGSPSPSANSMMG---PSPGPPSAG 53
QY 109 --TKAGIGGP-----AAAGAGCSPPRKYQAV--LPITQGSILVAAAKEPTPW 152
DB 54 HPITPQGGGATPQDNNHQMHRPMESNHEKMSDDPRYNQMGKMRSGHAGCPPPSPW 113
QY 153 AGDKGGAAPATASDPAGPPPLPGLPPPLAPATAGTLAASBGRKMSRKSPILGGGG 212

```

OY 487 RDNPSD 493
Db 1126 KPPPTPE 1132

RESULT 5
US-09-370-838-216
; Sequence 216, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ. ID NOS: 289
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ. ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-216

Query Match 5.44; Score 174; DB 4; Length 527;
Best Local Similarity 23.14; Pred. No. 0.00034;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

OY 26 ERRA--ALGPPDEPDAE-AHFLPRHRLK-----EPGPPLASOGGSPRSP 71
Db 8 EKAENALGSPPEEKVLENGELTPRREKALENGELSPPEAGEVIVNGILTPKSE 67
OY 72 AGCGGKRGILLPAGA-----APGOE--ESMGGSVPL--PCP-----PP 107
Db 68 DYSSENG-GLRFPNTRPPTGWRAPGPKWETPESMGVAPITIGEPAPETSLERAPPS 126
OY 108 ATGAGIGEPAAAGACSPRYOAVLP-----IOTGSLVAAKEPTWAG--DKG-G 158
Db 127 AVSSRRNGETACPLG--PAPKNGLEPETERRAPETGAPRA--PGAGRLDGG 179
OY 159 AASPAATSDPAG-----PPPLRPGP----- 181
Db 180 GRAVGTGTAPGGGPGGVDAKAGWVNTNRPOPPLPPPEAPRLEPAPRAREV 239
OY 182 ----PLAPTATAGTLAEGRW---KSMRSPGSGGSGASSQAACLQILLLODLI 233
Db 240 APSEPGAPDSRAGDGTALSGDDPPKPKRGP-----EMRPLFLDLG 282
OY 234 EQQOQOQAKKEIKSERDILLARIEMRMOLVKKONEKERHKLFOGYET----- 288
Db 283 PPGNSHQIKAR-----LSRLSLALPLTLTPPGGPRRPPMEGADAGAAG 330
OY 289 -----FERETELSEKIKLECOPELSTGTLPPKPSGCRSGKHRRKSPF- 335
Db 331 BAGGAGAPPAEDEDGEDEDEDE-----EDEEAAPGA-AAGPRGGRARRAIVP 379
OY 336 ----GSTERKTPYKLAPEFSKWTPTPKHSPTKEBPGSLSETVCKRELRS----- 383
Db 380 VVVSASADAAARLRLG-----LKSFGADEPED-----SLEKKRKNVSHGQVTV 426
OY 384 ----QETPEK-----PRSSVD-----TPRLSTPQK--PSTHPKKAASSET-- 420
Db 427 YLFDOETPTNELSVQAPPEGDPTDPSTPAPPTPPHPATPGDGFPSS--NDSGFGSPEMA 483
OY 421 EDLPYLSTEMYLCRMHQPPSPPLPLRESSPKKEETVARCMLSPSSVAGESVLAIVPMRD 480
Db 484 EDEPLT-----PPGP-PL-----CFSRSV--SPALETTPG-- 511

OY 481 HSYEPLRDPN--PSDILEN 497
Db 512 ----PPARAPDARACPEVN 527

RESULT 6
US-08-976-255-11
; Sequence 11, Application US/08976255
; Patent No. 6136581
; GENERAL INFORMATION:
; APPLICANT: Jono, Keith E.
; APPLICANT: Plozman, Gregory
; TITLE OF INVENTION: KINASE GENES AND USES
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FASTSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,255
; FILING DATE: No. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/031,675
; FILING DATE: No. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; ;
; INFORMATION FOR SEQ. ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1384 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-976-255-11

Query Match 5.38; Score 173; DB 4; Length 1384;

Best Local Similarity 20.94; Pred. No. 0.0013;
Matches 164; Conservative 79; Mismatches 265; Indels 278; Gaps 42;

OY 10 AAAAPAGGNPEQLDYERAAAL-----GCPDEPDAE-AHFLPRHRLKPE----- 55
Db 561 APSPATADDDDDSDGTAASLAMEPLLGHGPPVDVPMGGRDHY--FRSLARPLCPSPS 619
OY 56 ----GPIASSOGGS-----PA-----PSPACCG-----KGRGLLPAGA 87
Db 620 PSPSAGLSLAEGGADBDGVAFCAPFEDPLGSPGSSGAPPLPLTGBDELEBVG 679
OY 88 APGOQESWGGSVPL-----PCPPATKAGIGGEPAAGA--GC-----SPRP 129
Db 680 RRAAQRGHRNSVANSNNSGRCPE-----SMDPVSAAGHAECGCPSPKQTPRASPE 731
OY 130 KY--QAVLPQOTGSLVAAKEP-----TP-W-----AGDKG 158
Db 732 GTFEPBLGLQ-----AASAOEPCCEPLPHLCSAOGIAPAPCLVITTSWTETASSGGDHPQ 787


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/574,959A
FILING DATE: 19-DEC-95
ATTORNEY/AGENT INFORMATION:
NAME: Mandagouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-574-959A-7

```

```

Query Match      6.2%; Score 199.5; DB 2; Length 1135;
Best Local Similarity 21.9%; Pred. No. 1.7e-05;
Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22;

```

```

QY 38 PGAAEAHFLPRHKK-----LKEGPPPLASSGSGSPAPSPAGCGGKRGILLPAGAAPG 90
Db 644 PTPAPVPLRPHRPSGPHRSILRAPCPQMA-----PCPQAPCPSPAG-----PMSAGP 692
QY 91 QOESWCG-----SVLPDPP-----PATKQAGIGEPAAAGAGCS-----126
Db 693 VPSEPTSTTANLLGLLSRPSVCPRLPGPNHAGSNEDPILAPSGTPPTTPDET 752
QY 127 ----PRPKY-----QAVLPQIGSLVAAKE-----PTPWAGDKG 157
Db 753 GGVPRPAFVHYDKERASDVETLSLSDSDSVIVPEGLPLPLPPPPSGATPPPIAPATGP 812
QY 158 GAASPAATASD-----PAGPPPLPLPGPPP--LAPTATAGTLAASGRKMSMRKSPGLG 209
Db 813 PTASPPVPAKEEPEELPAPGPLPPPPPPPPVPGVXLPPLQVPGS-----TP--G 863
QY 210 GGGSGASSQAACLKQILLDLLEQOQOOLQAKETELKSERDTLLARIRMRMQ 269
Db 864 GGGPALDE-----DLVIVININSSDEEEEGEEDDEEEEEE-----EEDEEE 912
QY 270 LVKKDNEKRNKLFQGYETE-----EREETELSEKIKLECPLETSQTILPPKPS 321
Db 913 EDFEEEDDEEYFEEEEEDEEEFEFEFEELBEEDEDEDEEELEEV---EDLE 969
QY 322 CGRSGKHKRKSPPGSTERKTPVKKLAPESKVTKTPKHSPIKEEPCGSLSETVCKREL 381
Db 970 FGTAQ-----GEVEEGAPPPPLPALPPSPSPKVPQPEPEPGLILEV-----1014
QY 382 RSQETPEKPRSSVDTPPRLS-----TPQKGPSTHPEKKAFFSSEIDLPYL 426
Db 1015 --EEPGTEERGADTATPLAPALPQSGEVEERGESAPAGP---PPQELVEEPPSXPPPL 1069
QY 427 STEMYLCRMHQPPSPPLPRESSPKKEETVACMLPSSVAGETSVLAVPSWRDHSVEPL 486
Db 1070 LEEETEDGSKVOPPETPAEE--METETEALQEKEDDTAAMLA--DFIDCPPDE 1125
QY 487 RDPNPSD 493
Db 1126 KPPPE 1132

```

```

RESULT 4
US-09-357-014-7
Sequence 7, Application US/09357014
Patent No. 6291645
GENERAL INFORMATION:
APPLICANT: Jaekyoon Shin, Insil Jouny, Ratna K. Vadlamudi
and Jack L. Strominger
TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
AND USES THEREFOR
NUMBER OF SEQUENCES: 22

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,014
FILING DATE: 19-Jul-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/574,959
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mandagouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-357-014-7

```

```

Query Match      6.2%; Score 199.5; DB 4; Length 1135;
Best Local Similarity 21.9%; Pred. No. 1.7e-05;
Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22;

```

```

QY 38 PGAAEAHFLPRHKK-----LKEGPPPLASSGSGSPAPSPAGCGGKRGILLPAGAAPG 90
Db 644 PTPAPVPLRPHRPSGPHRSILRAPCPQMA-----PCPQAPCPSPAG-----PMSAGP 692
QY 91 QOESWCG-----SVLPDPP-----PATKQAGIGEPAAAGAGCS-----126
Db 693 VPSEPTSTTANLLGLLSRPSVCPRLPGPNHAGSNEDPILAPSGTPPTTPDET 752
QY 127 ----PRPKY-----QAVLPQIGSLVAAKE-----PTPWAGDKG 157
Db 753 GGVPRPAFVHYDKERASDVETLSLSDSDSVIVPEGLPLPLPPPPSGATPPPIAPATGP 812
QY 158 GAASPAATASD-----PAGPPPLPLPGPPP--LAPTATAGTLAASGRKMSMRKSPGLG 209
Db 813 PTASPPVPAKEEPEELPAPGPLPPPPPPPPVPGVXLPPLQVPGS-----TP--G 863
QY 210 GGGSGASSQAACLKQILLDLLEQOQOOLQAKETELKSERDTLLARIRMRMQ 269
Db 864 GGGPALDE-----DLVIVININSSDEEEEGEEDDEEEEEE-----EEDEEE 912
QY 270 LVKKDNEKRNKLFQGYETE-----EREETELSEKIKLECPLETSQTILPPKPS 321
Db 913 EDFEEEDDEEYFEEEEEDEEEFEFEFEELBEEDEDEDEEELEEV---EDLE 969
QY 322 CGRSGKHKRKSPPGSTERKTPVKKLAPESKVTKTPKHSPIKEEPCGSLSETVCKREL 381
Db 970 FGTAQ-----GEVEEGAPPPPLPALPPSPSPKVPQPEPEPGLILEV-----1014
QY 382 RSQETPEKPRSSVDTPPRLS-----TPQKGPSTHPEKKAFFSSEIDLPYL 426
Db 1015 --EEPGTEERGADTATPLAPALPQSGEVEERGESAPAGP---PPQELVEEPPSXPPPL 1069
QY 427 STEMYLCRMHQPPSPPLPRESSPKKEETVACMLPSSVAGETSVLAVPSWRDHSVEPL 486
Db 1070 LEEETEDGSKVOPPETPAEE--METETEALQEKEDDTAAMLA--DFIDCPPDE 1125

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:11 ; Search time 28 Seconds
(without alignments)
645.203 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238
Sequence: 1 MTMRSAVFKAAPAGAGNPE.....RSRCRLTIQKQTPHRTCRK 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	199.5	6.2	905	US-08-574-959A-9	Sequence 9, Appl1
2	199.5	6.2	905	US-08-357-014-9	Sequence 9, Appl1
3	199.5	6.2	1135	US-08-574-959A-7	Sequence 7, Appl1
4	199.5	6.2	1135	US-08-357-014-7	Sequence 7, Appl1
5	174	5.4	527	US-09-370-838-216	Sequence 216, App
6	173	5.3	1384	US-08-976-255-11	Sequence 11, Appl
7	171	5.3	543	US-09-535-008-63	Sequence 63, Appl
8	171	5.3	577	US-09-535-008-61	Sequence 61, Appl
9	171	5.3	1646	US-09-535-008-67	Sequence 67, Appl
10	171	5.3	1647	US-09-535-008-2	Sequence 2, Appl1
11	171	5.3	1649	US-09-535-008-75	Sequence 75, Appl
12	171	5.3	1650	US-09-535-008-71	Sequence 71, Appl
13	171	5.3	1678	US-09-535-008-69	Sequence 69, Appl
14	171	5.3	1679	US-09-535-008-65	Sequence 65, Appl
15	171	5.3	1681	US-09-535-008-77	Sequence 77, Appl
16	171	5.3	1682	US-09-535-008-73	Sequence 73, Appl
17	169.5	5.2	3969	US-08-061-376-5	Sequence 5, Appl1
18	164	5.1	1618	US-07-853-913-4	Sequence 4, Appl1
19	162	5.0	1593	US-08-628-829-4	Sequence 4, Appl1
20	160.5	5.0	1185	US-09-041-886-23	Sequence 4, Appl1
21	160	4.9	1137	US-08-369-043-2	Sequence 23, Appl
22	158.5	4.9	504	US-09-219-849-3	Sequence 3, Appl1
23	158.5	4.9	561	US-08-642-255-52	Sequence 52, Appl
24	158	4.8	960	US-09-219-849-5	Sequence 5, Appl1
25	154.5	4.8	1493	US-09-423-890-8	Sequence 8, Appl1
26	154	4.8	720	US-09-219-849-4	Sequence 4, Appl1
27	154	4.8	777	US-08-642-255-53	Sequence 53, Appl

28	152.5	4.7	234	US-08-642-255-51	Sequence 51, Appl
29	152.5	4.7	1064	US-08-642-255-62	Sequence 62, Appl
30	151.5	4.7	2414	US-08-227-536-2	Sequence 2, Appl1
31	151.5	4.7	2414	PCT-US95-04682-2	Sequence 2, Appl1
32	151	4.7	1060	US-08-931-820-3	Sequence 3, Appl1
33	151	4.7	1418	US-08-963-825-20	Sequence 20, Appl
34	151	4.7	1418	US-09-010-999-1	Sequence 1, Appl1
35	151	4.7	1418	US-09-500-811-20	Sequence 20, Appl
36	151	4.7	1418	US-09-570-573-20	Sequence 20, Appl
37	151	4.7	1418	US-09-548-608-20	Sequence 20, Appl
38	150.5	4.6	1274	US-09-095-443-2	Sequence 2, Appl1
39	149.5	4.6	529	US-09-247-806-2	Sequence 2, Appl1
40	149.5	4.6	595	US-08-425-069-4	Sequence 4, Appl1
41	149.5	4.6	595	US-08-317-844B-4	Sequence 4, Appl1
42	149.5	4.6	902	US-08-396-479B-6	Sequence 6, Appl1
43	149.5	4.6	902	US-08-818-823-6	Sequence 6, Appl1
44	149.5	4.6	2441	US-08-194-468-2	Sequence 2, Appl1
45	149.5	4.6	2441	US-08-961-739-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-574-959A-9
: Sequence 9, Application US/08574959A
: Patent No. 5962224
:
GENERAL INFORMATION:
: APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
: APPLICANT: and Jack U. Strominger
: TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
: TITLE OF INVENTION: AND USES THEREFOR
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESS: LAHIVE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/574,959A
: FILING DATE: 19-DEC-95
: ATTORNEY/AGENT INFORMATION:
: NAME: Mandragouras, Amy E.
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: DEN-008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 905 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
US-08-574-959A-9
:
Query Match 6.2%; Score 199.5; DB 2; Length 905;
Best Local Similarity 21.9%; Pred. No. 1.3e-05;
Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22;
:
Oy 38 PGAAEAFLPRHRK-----LKPPGPLSSQSGSPAPSPAGCGGKRGILLPAGAPG 90
Db 414 PTPAPVPLLRPHRSRHSILRAFCPQMA-----PCPOAPCPSSAG-----PMSAGP 462
Oy 91 QOESMCG-----SVPLPCPP-----PATKACIGGEPAAGAGCS----- 126
```


; Sequence 4, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenjin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DVL
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019-4

Query Match 5.3%; Score 170.5; DB 9; Length 1610;
Best Local Similarity 22.1%; Pred. No. 0.33;
Matches 134; Conservative 51; Mismatches 203; Indels 217; Gaps 30;

QY 10 AAAAPAGNP--EORLDYERAA-----ALGAPDEDPG---AAEAHFLPRHRLKKEPG 56
DB 550 AGALPGLREPLMEHRYLEEAAREBQATILAKAPSEFTALRLPASGTHLAPGSHSLEHD 609
QY 57 PPLASQGGSPAPSPAGCCGKGRGLLPAGAAP-----GQGESWGSVPLPCPP 107
DB 610 SP-----STPRSSACGEAQRLPSPAGAPIRDMGHQSGKQLPSTGHH-PGTAQPE 662
QY 108 ATKQAGIGGEPA-----AAGACSPRYKQAVLPQTGSL--VAAAKE-----PTW 152
DB 663 RSPSPWQGPAPFCHPKQGSAPQEGCSFHP--AVAPCPGSPGSCKEAPLVSSPF 719
QY 153 ACDKGASPAATASDP---AGPPPLPPLPPLPPLATAGTLA--ASEGRKMSKSP 207
DB 720 LQOPAPAPAP--AKASPLDSKMGPDISLPGRPKPGPCSPGASASQVSSILRVSS 778
QY 208 GGGGSGASGSAACQKILLQLDLIEQQOOLAKKEIEELKSERDTLLARIEMERR 267
DB 779 QYGTGEPGSLDA-----EGWTQEAEDLSDSTPTL----- 807
QY 268 MQLVKKDNEKERHKLQ-----GYET-----EEREE 293
DB 808 ---ORPOEQATMRKFSLGGRGYAGVAGYGTFAFGDAGMLGCPMMARIANAVSOS 863
QY 294 TELSEKILECOPELSETSQTLPPKPPSCGRSGKHRRKSPFGSTERKTPYKKLAPFSK 353
DB 864 EEDDEARAESEEQOEARAESPLPOVSAR-----PV-----PEVGR 901
QY 354 VKTTPKHSPIKEEPCGSLSETVCKREL---RSQEMPEKRRSSVDTPPLSTPOKGPST 409
DB 902 APTRS--SPEPTPWEDIGV--LYQIRDSGDAEADTISLDISVD--PAYLNL----- 951
QY 410 HPKEKAPSETEDLPYLSSTEMYLGRWHQPPSPPLPRESSPKKETVARCIMPSSVAGE 469
DB 952 ---SDLYDIKYLPEFEMI---FRKVPKSAQ--EPSPMAEEELAE----- 989
QY 470 TSVLAVPSMRHSDVEPLRDNPSPDLLENLDSVFSKRHAKLELDEK----- 515
DB 990 ---FPPEPTW-----FWPGEL-----GPHAGLEITESEEDVDALLAEAAVG 1026
QY 516 RRRKW 520
DB 1027 RRRKW 1031

Search completed: July 14, 2003, 12:00:17
Job time : 58 secs

QY 597 RCRLEIOKQTPHRTCR 613
| : : : :
Db 602 EMRSVLKRVGSPRKRR 618

RESULT 13
US-09-824-574-4
; Sequence 4, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Molanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: Arip4 Gene and protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-574-4

Query Match 5.3%; Score 171; DB 9; Length 1647;
Best Local Similarity 19.9%; Pred. No. 0.32;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKPPGPPPLASSOGSPAPSPAGCGGKGRGLL--PAGAPGQOEESWGSVPLPCPPA- 108
| :
Db 1 MSPDPPL---GTPRPSPSPGPPSGAMLGPPSPGSPGSAHSMWG---PSPGPPSAG 53
QY 109 --TKAGIGSEP-----AAAGACSPRPKYQAV--LPITGSLVAAKETPM 152
| :
Db 54 HPIPTGPGGYPDNDNHQMKPMESMHEKMSDDPRNOMKMGMRSGHAGMGPPSPM 113
QY 153 AGDKGAASPAATASDPAGPPPLPLPGPPPLATATAGTLAASEGRKMSRKSPDLGGG 212
| :
Db 114 DMSOGYPSPLGSEHASSVPASGSSG-----QMSGP--GGAPL 154
QY 213 SGASSQAACIKQILLDLLEQOQOOLAKKEIEELKSE--RDTLLARIEMERMOL 270
| :
Db 155 DGADPOA-----LGQNRGPTPPNQNQLHQLRAQIMAYKMLARGOPLPDHQM 202
QY 271 VKNDNEKHKLLQGYETERETELSEKIKLEQPELSTSTLPPKPPSCGRSGK-- 328
| :
Db 203 A-----VQGRKP-----MPMQOQMPTLLPPSVSATGPGPG 235
QY 329 -----HKRSPGSTERKTPYVKLAPEFSKYTKTPKHSPIKEEPCGSLSET 375
| :
Db 236 RCGPGRCGAPRPNYSRPHGGMNMPRPGSGVPRPGPGQ---PRGPRPRPMBGRMANA 292
QY 376 VCRBELSDETPKPPNSVDTPRLSTPOKGPSTHPEKAFSEIETDLPYLSTTEMYLCR 435
| :
Db 293 AAFSTPQKLIPOQPRGPRPAPPAVPMPAPOTQSPGQPAQAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESPPKKEETVARCL-----MPSVAETSVLAPSN 478
| :
Db 346 -HKQSRITPTQCRGLDPEILLQERFYLQANRIARHIOLEMLPGSLADLTAKA--- 401
QY 479 RDHSVEPLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQRIOREORILQIOL 536
| :
Db 402 --TELKALRLNFORQRLQ--EYVVCMKRDTALFALNAKAYKRSKQSLSEARITELK 458
QY 537 RM---YKKGIQSEPEVYTSFPEPDD 560
| :
Db 459 QOKIEQERRRKHQETLNSIIOHARD 485

RESULT 14
US-10-307-019-1

; Sequence 1, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DVI
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019-1

Query Match 5.3%; Score 170.5; DB 9; Length 1351;
Best Local Similarity 22.1%; Pred. No. 0.27;
Matches 134; Conservative 51; Mismatches 203; Indels 217; Gaps 30;

QY 10 AAAAPAGGNP--EQRIDYERA-----ALGGPEDEPG---AAEAHPLRHRKLEKPG 56
| :
Db 291 AGALPGIREPLMHRVLEEFNAAREQATLLAKAPSETALRLPASGTHLAPGSHSLEND 350
QY 57 PLLASSOGSPAPSPAGCGGKGRGLLPAGAP-----GQOEESWGSVPLPCPP 107
| :
Db 351 SP-----STPRSSACGAEQRLPSAPSGAGATIRMGHPGSKQLPSTGCH--PGRAPQ 403
QY 108 ATQAGIGSEP-----AAGACSPRPKYQAVLPITGSL--VAAKE-----PTM 152
| :
Db 404 RSPDSPWGPAPFCHPKQGSAPQEGCSBP--AAAPCPGSPFPSCKEADLVPSPE 460
QY 153 AGDKGAASPAATASDP-----AGPPPLPLPGPPPLATATAGTLA--ASEGRKSMKRSPL 207
| :
Db 461 LGPQAPAP--AAASPLPLDSKMGPGDISLGRKRPCCSPGASQAASSQVSSLVGSS 519
QY 208 GGGGSGASSQAACIKQILLDLLEQOQOOLAKKEIEELKSERDTLLARIEMERR 267
| :
Db 520 QVTEPGPSLDA-----EGMTQEAEDLSDSTPL----- 548
QY 268 MQLVKDNEKHKLLQ-----GYET-----BEREE 293
| :
Db 549 ---ORPEQATMKRKSLSGSGGYAGVAGTGTEAFGADAGMLGQPMARIAMAVQSE 604
QY 294 TELSEKIKLEQPELSTSTLPPKPPSCGRSGKHKRKSPPGSTERKTPYVKLAPEFSK 353
| :
Db 605 EEEQEFKRAASQSEBQDEAARLPOYSK-----PV-----PEYGR 642
QY 354 VKTKTPKHSPIKEEPCGSLSETVCKREL---RSQETPEKPPNSVDTPRLSTPOKGPST 409
| :
Db 643 APTRS--SPEETPMEDIGQVS--LVQIRDLSDADAADTIISLDISEVD--PALNL----- 692
QY 410 HPEKKAFFSSEIETDLPYLSTTEMYLCRWHOPRPPPLRESPPKKEETVARCLMPPSSVAGE 469
| :
Db 693 -----SOLYIKIYLPFEFKT--FRKVPKSAQ--EPSPMAEEELAE----- 730
QY 470 TSVLAVPSMDHSVEPLRDPNPSDLLENLDSVFSKRAKLELDER----- 515
| :
Db 731 ---FPEPTW-----FWPEEL-----GHAGLEITESESDVDALLAAVAG 767
QY 516 RKRKW 520
| :
Db 768 RKRKW 772

RESULT 15
US-10-307-019-4

```
Qy      481 HSEVPLRDPN--PSDLEN 497
          | | | : | : : ||
Db      512 ---PPARAPDARPAQPVEN 527
```

```

RESULT 11
US-09-919-039-278
; Sequence 278, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTUREDRESS
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 278
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
; US-09-919-039-278

```

```

RESULT 12
US-10-171-384-1
: Sequence 1, Application US/10171384
: Publication No. US20030031680A1
: GENERAL INFORMATION:
: APPLICANT: The Brigham & Women's Hospital, Inc.
: APPLICANT: Belzer, David
: APPLICANT: Herion, Bruce
: APPLICANT: Rao, Cherie
: TITLE OF INVENTION: p33 Binding Protein-Related
: FILE REFERENCE: 81994/275368
: CURRENT APPLICATION NUMBER: US/10/171,384
: CURRENT FILING DATE: 2002-06-14
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 824
: TYPE: PRT
: ORGANISM: mouse
: US-10-171-384-1

```


RESULT 9
US-10-144-649A-216
; Sequence 216, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-144-649A-216

Query Match 5.4%; Score 174; DB 9; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA---ALGPPDEPGAEE-AHFPRHRKX-----BPGPIASSOGGSPAPSP 71
DB 8 EKAENGALGSPREKREKVLNGLTPPRREKALENGELSPAGKVLVNGSLTPPKSE 67
QY 72 ACCGGKRGILLPAGA-----APGOE---ESWGSVPL-PCP-----PP 107
DB 68 DKVSENG-GLRFRNTERPETGPMRAPGWEKTPSMGAPRTIGEPAPETSLERAPAP 126
QY 108 ATQAGTIGEPAAAGACSPRKYQAVLP-----IQTGLVAAKEPPIPMAG--DKG-G 158
DB 127 AVVSSRNGGETAGPLG--PAPKNGTLEGTERRABETGAPARA-----PGAGRLDIGSG 179
QY 159 AASPATYASDPAG-----PPPLPLGPP-----PP 181
DB 180 GRAPVGTGTAPGGGPGSSVDKAGWYDNTRRPPPPPLPPPPPAQPRRLPPAPRPAP 239
QY 182 ----PLAPYATAGTLAASEGRW--KSMRKSPLGGGGSGASSQAACIKQILLQDLI 233
DB 240 APEGEGAPDSRAGGDTALSGDGDPPKPERKGP-----EMRPLFLDIG 282
QY 234 EQQOQOOLQAKKEIEELKSRDILLARIEMERRMOLVKKDKERHKLFGIET----- 288
DB 283 PPGGNSQILKAR-----LSRLSLALPPLTLTPPGGPRRRPMEGADAGAGG 330
QY 289 -----ERETELSEKIKLECOBELSETQTLPPKPFSCGSGKGRKSPF- 335
DB 331 EAGGAGAPGAEDGEDDEDE-----EDEEAAPGA--AAGPRGPARAPAPV 379
QY 336 -----GSTERTPVKTLAEPFSVKTKTPRHSPIKEEPCGSLSETVCKRELRS----- 383
DB 380 VVVSADADAARPLRGL-----LKSPPGADEPD-----SELERKRMVSHFDVTV 426
QY 384 -----QETPEK-----PRSSVD-----TPRLSTPOK--PSIHREKAFSSSI- 420
DB 427 YLFDETPNELSVQAPPGGDTDPSTPPAPPTPHATPGDGPFS--NDSGGSGFEWA 483
QY 421 EDLPYLTSTEMYLCRWHOPPPSPPLRESSPKKEETVARCLMPSSVAGETSVLAVPSMRD 480
DB 484 EDFPPL-----PPPGP-PL-----CFSRFSV--SPALETG--- 511
QY 481 HSYVEPLRDRN--PSDILEN 497
DB 512 ---PARAPDARAPGAVEN 527
RESULT 10

US-09-738-973-216
; Sequence 216, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-216

Query Match 5.4%; Score 174; DB 10; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA---ALGPPDEPGAEE-AHFPRHRKX-----BPGPIASSOGGSPAPSP 71
DB 8 EKAENGALGSPREKREKVLNGLTPPRREKALENGELSPAGKVLVNGSLTPPKSE 67
QY 72 ACCGGKRGILLPAGA-----APGOE---ESWGSVPL-PCP-----PP 107
DB 68 DKVSENG-GLRFRNTERPETGPMRAPGWEKTPSMGAPRTIGEPAPETSLERAPAP 126
QY 108 ATQAGTIGEPAAAGACSPRKYQAVLP-----IQTGLVAAKEPPIPMAG--DKG-G 158
DB 127 AVVSSRNGGETAGPLG--PAPKNGTLEGTERRABETGAPARA-----PGAGRLDIGSG 179
QY 159 AASPATYASDPAG-----PPPLPLGPP-----PP 181
DB 180 GRAPVGTGTAPGGGPGSSVDKAGWYDNTRRPPPPPLPPPPPAQPRRLPPAPRPAP 239
QY 182 ----PLAPYATAGTLAASEGRW--KSMRKSPLGGGGSGASSQAACIKQILLQDLI 233
DB 240 APEGEGAPDSRAGGDTALSGDGDPPKPERKGP-----EMRPLFLDIG 282
QY 234 EQQOQOOLQAKKEIEELKSRDILLARIEMERRMOLVKKDKERHKLFGIET----- 288
DB 283 PPGGNSQILKAR-----LSRLSLALPPLTLTPPGGPRRRPMEGADAGAGG 330
QY 289 -----ERETELSEKIKLECOBELSETQTLPPKPFSCGSGKGRKSPF- 335
DB 331 EAGGAGAPGAEDGEDDEDE-----EDEEAAPGA--AAGPRGPARAPAPV 379
QY 336 -----GSTERTPVKTLAEPFSVKTKTPRHSPIKEEPCGSLSETVCKRELRS----- 383
DB 380 VVVSADADAARPLRGL-----LKSPPGADEPD-----SELERKRMVSHFDVTV 426
QY 384 -----QETPEK-----PRSSVD-----TPRLSTPOK--PSIHREKAFSSSI- 420
DB 427 YLFDETPNELSVQAPPGGDTDPSTPPAPPTPHATPGDGPFS--NDSGGSGFEWA 483
QY 421 EDLPYLTSTEMYLCRWHOPPPSPPLRESSPKKEETVARCLMPSSVAGETSVLAVPSMRD 480
DB 484 EDFPPL-----PPPGP-PL-----CFSRFSV--SPALETG--- 511

367 -----BPCSLSEYVCKRELRSOETPE-KPRSSVDPDP----- 398
Db 991 KPVPPVPPPHLOPBEEDVSO-----QSGSGPRGKSRSRV--PRAKEAKAPFPAPF 1041
QY 399 -----RLSTPQKG-----PSTHPEKAFSSIEDLPLSTTEMYLCRMHOPPSPPLPL 446
Db 1042 PTEGOSYRLS-PHAGHRLPSHPREVYIKTSTRAD-PLFSTY-----PPGHPLPL 1088

RESULT 7
US-09-862-027-80
; Sequence 80, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodges, Martin R.
; TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345,473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 1265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-027-80

Query Match 5.4%; Score 174.5; DB 10; Length 1265;
Best Local Similarity 23.6%; Pred. No. 0.15;
Matches 134; Conservative 51; Mismatches 201; Indels 183; Gaps 28;

QY 4 RSNVFAKAAAPAGNPPQRLDYERAAALGPEDEPGAENFLPRHKKLKPPGPLA--- 60
Db 759 RSSLFRTKQASLLHTSRLSLNRSLSGSGSPGS-----PTHSHSLSPRSPTQYR 812
QY 61 -----SSGGSPA-----PSPAGCGKRGILLPAGAPGOOE-----SMWG 98
Db 813 VTPNAVHVGNSQSSSPSSVSPSPAGSHTRPSSL--HGLAPKIQORSPRRKSAG 870
QY 99 SVPL-----PCPPATKQAGIGGEPAAGACSPRPKYQAVLPDIQTGSLVAAAKPTPW 152
Db 871 SIPLSLAHTPSPPPPT-----ASPO-----RSPSPL 897
QY 153 AGDGGAASPAATSDAPGP-PLPLPGPPPLAP-----TATAGTLASEGRKMSMR 203
Db 898 SGHVAQAFTKYLHLSPLGRLSRKSAEPSPPLKRVQSAEKLAAALASEKKLATSR 957
QY 204 KSPAGGGGSGASSQAACLKQILLLODLLEOQ--OOOLQAKEIEELKSERDYL----- 257
Db 958 KH-----SIDLPHSELKELPREVSPLEVGARSLSGKA 994
QY 258 -----LARIEMERMOLVKKDNEKERHKLFGYEIEE-RETELSEKI 300
Db 995 LPGKGVLPAPSRALGTLRODRAERRESLOKOEAIREVDS--SEDDTEEGENSGAOEL 1052
QY 301 KLECQPELSESOFLPPKPSGSGSGKHKRSPFGSTERTKPYKTLAPESXYKTKTPK 360
Db 1053 SLAPHEV--SOSVAPK--GAGESG---EDFPFS---RGP--RSLGPMVPSLTLGTL 1099
QY 361 HSPKEEPCGSLSTVCKRELRSOETPEKPRSSVDPPLRLSTPQKGFTH--PKK---KAF 416
Db 1100 GPPMESPSPG-----HRRIGSPQATEEAASSSSAGPNLG--QSGADTLPPEGCMAQ 1151
QY 417 SSEIEDLPLSTTEMYLCRMHOPPSPPLPRESSPKKEETVYARCLMPSSVAGETSVLAVP 476
Db 1152 HLHTQALTAIS-----PSTSGTLPTS-----SCSPSSTSGK---LSMW 1187
QY 477 SMRD-----HSVEPLRDPNPSDLEND 500
Db 1188 SMKSLIEGPDRASSPRKATYAGGLANLQD 1216

RESULT 8
US-09-854-133-216
; Sequence 216, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-216

Query Match 5.4%; Score 174; DB 9; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA---ALGPEDEPGAEE-AHFLPNRRIK-----EPGLASSGGSPAPSP 71
Db 8 EKAENALGSPEREERVELNGELTPPRREKALENGELRSPEAGEKVELVNGTLTPKSE 67
QY 72 AGCGKRGILLPAGA-----APGOE-----ESMGSGVPL--PCP-----PP 107
Db 68 DKVSENG-GLEPNTERTPEPTGWPAPGPEWKEPTESMGPAVTTGEPAPETSLERAPAP 126
QY 108 ATKQAGIGGEPAAGACSPRPKYQAVLP-----IOTGSLVAAAKEPTPWAG--DKG-G 158
Db 127 AVVSSRRNGETAPPLG--PAPKNGTLEPTGERAPPELGGAPRA-----PGAGLIDLGSG 179
QY 159 AASPAATASDPAG-----PPPLPGPP----- 181
Db 180 GRAVGTGTAVGGGPGSGVDKAKAGVNDTRPQPPPLPPEPEAPRLBAPRAPREV 239
QY 182 -----PLAPRTAGTLAASBGRW---KSMRSPGSGGSGASSQAACLKQILLLODLI 233
Db 240 APGEPPAPDSRAGDRLASDGDPPKPERKGP-----EMRFLFDLG 282
QY 234 EDOOQOLOAKEKEIEELKSERDTLLARIEMERMOLVKKDNEKERHKLFGYEET----- 288
Db 283 PPGNSRQIKAR-----LSRLSLALPLTLTPPGPGRPRRPMEGADAGAAG 330
QY 289 -----ERRETELSEKIKLECQPLSETSQTLPPKPPSCGSGSGKHKRSPF- 335
Db 331 EAGGAGAPPAEEDGEDEDEE-----EDEMAAGA--AAGPRGGRARAAPVP 379
QY 336 -----GTERKTPPKTLAPESKVTTPKPSKEEPCGSLSTVCKRELRS----- 383
Db 380 VVVSADADARPLRGL-----LKSPPRADPEP-----SELEKRRKRVSHGDTYV 426
QY 384 -----QETPEK-----PRSSVD-----TPRLSTPQKG--PSTHPEKAFSSEI--- 420
Db 427 YLFOETPTNLSVQAPPEGDTDPSTPPAPTPPHPATPGGFPSS--NDSGFGGSPFMA 483
QY 421 EDLPLSTTEMYLCRMHOPPSPPLPRESSPKKEETVYARCLMPSSVAGETSVLAVP 480
Db 484 EDLPPL-----PPGP--PL-----CPSRFSV--SPALETGPG--- 511
QY 481 HSVEPLRDPN--PSDLEND 497
Db 512 ---PARAPDARPAVENV 527

QY 421 EDLPYLSTTEMYLCRWHPSPPLPRESSPKKEEYVA 458
Db 241 EDLPYLSTTEMYLCRWHPSPPLPRESSPKKEEYVA 278

RESULT 2

US-09-893-519A-14
; Sequence 14, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Farid
; APPLICANT: ZHU, Shunao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/16548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/CA72189
; DATABASE ENTRY DATE: 1997-06-25
; RELEVANT RESIDUES: (1)...(1023)
US-09-893-519A-14

Query Match 6.4%; Score 206; DB 9; Length 1023;

Best Local Similarity 22.1%; Pred. No. 0.0025;

Matches 152; Conservative 72; Mismatches 264; Indels 200; Gaps 28;

QY 11 AAAPA-----GNEQRLDYERRAALGPEDEPGAAEHPLPRHKLKEGCPPLASSOGGS 66
Db 208 AAAPAASLVNNGPAALPLPKPAAPGVITQTPFVGAAAPAPAAASPPAAPAPAA 267
QY 67 PAPSAPCGCGKGRGLLPAG---AAP-----GGQERSMGSGVPLPPEPATQAGTGE 117
Db 268 PPPPPPPATLAPPGHPAGPPPAAPVPPPAANGSAGAAPAPAP-PAAGGAGVSGQ 326
QY 118 PAAAGCGSPRRYQAVLPIQTSLSVAAKEPTPMAGDKGASPAATAS-----167
Db 327 PGGGAATAAPAPCVKAKSPK---VQAA-----PPAAQTLAASGPASTASMTIGPTMG 379
QY 168 ---DPAGPPPLPPEPPLAPVATAGTLAASGRMKSMKSPJGGGSGASSOACLK- 223
Db 380 ALPSPAVPP-PAAGTPTGLPKGAGAVT-----QSLSRTPATTSIGATLTPVLA 432
QY 224 -----OILLQDLIQQOQOQLOAKRK-----245
Db 433 RLDPONPTNIONFOLLPGMVLVRSENGOLLMIPQALAKOMQAQAHAPOTTMAPPAT 492

QY 246 -----EIEELKSEBDTLAR-----IERMRMOVLKDKNEKER-----279
Db 493 PPSAPVQVSTVQAPPTPIARQVPTPTTIKKVQSAQTIVQPSATLQRSVGVPQLVLGG 552
QY 280 -----HKLFGYETEEREETEELSEKIKLQCPLELSETQTLPPKPS 321
Db 553 AAOATASLGATAVQGTGPQRTVPGATTTSSAATETMENK-KCKNLS---TL-----IK 603
QY 322 CGRSGKHRRKSPFGSTERTKTPVKL-----APESKTKTPKPHSPIKEPCCS 371
Db 604 LASSGR-----QSTETAAVKELVONLIDGKIAEDFTSLYELMSSP---Qp-----649
QY 372 LSETVCKREL-----RSQETPEKPRSSVDPTRPL---STPOKGPSTHAK 412
Db 650 YLVPFLKSLPALRQLTPPSAARIQSQOQPPPTSQATATLAVLVSSVGR---TAGK 706
QY 413 EKAFSEIEDPLPLSTTE-MYLCRWHPSPPLPRESSPKKEEYVARCLMBSVAGERS 471
Db 707 TAAVTASALQPPVLSLQPTQVGVGKQGPPTPLVIOQ--PPKGAALIR--PPQVTLQGP 762
QY 472 VLAVPSMRBHSVPLRDPNPSDLLENLDSVFESKRAKL-----ELDEKRRKMDIORIR 526
Db 763 MVA-----LQPHRRIMLTTPQOVNISEESARILATNSLVGLTRSCNDEFTL 811
QY 527 EQRILRLQRLMYKKKGIOSEPEVTSF 554
Db 812 LQAPLQRILIEIKKHGITEHPDVVSY 839

RESULT 3

US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Accm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 14, 2003, 11:50:46 : Search time 56 Seconds
(Without alignments)
1276.460 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3338
Sequence: 1 MTRMSAVFKAAAPAGGNPE.....RSRCRLFIQKQTPHRTCRK 614

Scoring table: BIOSIM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1446	44.7	281	10	US-09-925-302-655
2	206	6.4	1023	9	US-09-893-519A-14
3	199	6.1	1479	10	US-09-864-761-36182
4	194	6.0	1479	9	US-10-156-761-11739
5	187.5	5.8	2701	9	US-10-171-311-83
6	176.5	5.5	2462	9	US-09-819-104A-5
7	174.5	5.4	1265	10	US-09-862-027-80
8	174	5.4	527	9	US-09-854-133-216
9	174	5.4	527	10	US-10-144-649A-216
10	174	5.4	527	10	US-09-738-973-216
11	172.5	5.3	767	9	US-09-919-039-278
12	172	5.3	824	9	US-10-171-384-1
13	171	5.3	1647	9	US-09-824-574-4
14	170.5	5.3	1351	9	US-10-307-019-1
15	170.5	5.3	1610	9	US-10-307-019-4
16	170.5	5.3	1665	10	US-09-858-664A-2
17	170.5	5.3	2596	9	US-10-307-019-6
18	170.5	5.3	2630	9	US-10-077-130-2
19	170.5	5.3	7968	9	US-10-077-130-5

20	168.5	5.2	1000	9	US-10-128-714-3305	Sequence 3305, Ap
21	165.5	5.1	802	10	US-09-823-240-2	Sequence 2, Appl1
22	165	5.1	1555	9	US-10-128-714-3298	Sequence 3298, Ap
23	165	5.1	1832	9	US-10-128-714-8298	Sequence 8298, Ap
24	164.5	5.1	758	10	US-09-904-987-5	Sequence 5, Appl1
25	164.5	5.1	3256	5	US-09-919-039-21	Sequence 21, Appl1
26	164.5	5.1	3256	10	US-09-919-172-98	Sequence 98, Appl1
27	164	5.1	1001	9	US-10-128-714-3240	Sequence 3240, Ap
28	164	5.1	1618	9	US-09-963-875-1	Sequence 1, Appl1
29	164	5.1	1618	9	US-10-136-891-2	Sequence 2, Appl1
30	164	5.1	1618	9	US-10-120-687-1	Sequence 1, Appl1
31	163.5	5.0	810	9	US-10-156-761-10081	Sequence 10081, A
32	162.5	5.0	538	9	US-09-976-740-43	Sequence 43, Appl1
33	162.5	5.0	538	12	US-10-023-529-43	Sequence 43, Appl1
34	162.5	5.0	538	12	US-10-023-523-43	Sequence 43, Appl1
35	162.5	5.0	1259	9	US-10-260-715-8	Sequence 8, Appl1
36	160.5	5.0	775	9	US-09-738-626-3773	Sequence 3773, Ap
37	159.5	4.9	1312	9	US-10-029-115-2	Sequence 2, Appl1
38	158.5	4.9	629	10	US-09-833-790-429	Sequence 429, App
39	158.5	4.9	1295	9	US-09-789-390-30	Sequence 30, Appl1
40	158.5	4.9	1295	9	US-09-789-390-32	Sequence 32, Appl1
41	158.5	4.9	1295	9	US-09-789-390-34	Sequence 34, Appl1
42	158.5	4.9	1295	9	US-09-789-390-37	Sequence 37, Appl1
43	158.5	4.9	1295	9	US-09-789-390-39	Sequence 39, Appl1
44	158	4.9	824	9	US-09-909-567B-53	Sequence 53, Appl1
45	158	4.9	912	9	US-09-291-417-26	Sequence 26, Appl1

ALIGNMENTS

RESULT 1
US-09-925-302-655
; Sequence 655, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 655
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-655

Query Match 44.7%; Score 1446; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.8e-70;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 PPLAPVATAGTTLAASGRKMSKRSPLGGGGGSGASSQAACLOILLDLLEQQOOL 240
DB 1 PPLAPVATAGTTLAASGRKMSKRSPLGGGGGSGASSQAACLOILLDLLEQQOOL 60
QY 241 QAKKEKELEKSERDITLARIERERMOVLVKKDNEERKTLQGYTEREFTELSEKI 300
DB 61 QAKKEKELEKSERDITLARIERERMOVLVKKDNEERKTLQGYTEREFTELSEKI 120
QY 301 KLEQPELSETQTLPPKPSGSGKGRKSPFGSTERTKTPVKKLAPFSVKTKTPK 360
DB 121 KLEQPELSETQTLPPKPSGSGKGRKSPFGSTERTKTPVKKLAPFSVKTKTPK 180
QY 361 HSPKPEPCGSLSETVKKRELRSQETPEKRRSYDTPPLSTQKGPSTHPKKAASSEI 420
DB 181 HSPKPEPCGSLSETVKKRELRSQETPEKRRSYDTPPLSTQKGPSTHPKKAASSEI 240

Tue Jul 15 10:16:07 2003

us-10-054-935-2.rni

Page 21

Db 441 GACACACCTCGGGCACCACCACCACTCCGGGCGCAAGCTCC-----394
 Oy 458 ALAArgCysLeuMetProSerSerValAlaIglYlnrHisSerValLeuAlaValProSer 477
 Db 393 -----ATTCCACCGGCATCATCCAGTCCTCCATCATCGTCATCGTCGTCCTCA 343
 Oy 478 TRPArgAspHisSerValGluProLeuArgAspProAspPro 491
 Db 342 TCATAGTCATATTCACCATCTCCACCGCGCGCGCTCCG 301

```
Search completed: July 14, 2003, 23:49:40
Job time : 4383 secs
```

ADDRESSEE: 812-5 Hirano
 STREET: Isshinden
 CITY: Tsu-city
 STATE: Mie-prefecture
 COUNTRY: JAPAN
 ZIP: 514-01

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Microsoft Windows 95
 SOFTWARE: Word Perfect 6.1

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/864,038A
 FILING DATE: May 28, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-184459
 FILING DATE: 15-July-1996

ATTORNEY/AGENT INFORMATION:
 NAME: C. Bruce Hamburg
 REGISTRATION NUMBER: 22,389
 REFERENCE/DOCKET NUMBER: F-5610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)986-2340
 TELEFAX: (212)953-7733

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3331
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Pinctada fucata
 CELL TYPE: mantle epithelial cell
 FEATURE: mRNA
 LOCATION: from 1 to 3331
 IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-2

Alignment Scores:
 Pred. No.: 0.0159 Length: 3331
 Score: 211.50 Matches: 128
 Percent Similarity: 34.83% Conservative: 58
 Best Local Similarity: 23.97% Mismatches: 235
 Query Match: 6.53% Indels: 116
 Gaps: 17

US-10-054-935-2 (1-614) x US-08-864-038A-2 (1-3331)

QY 6 AlaValPheLysAlaAlaAlaAlaProAlaGlyLysAsnProGluGlnArgLeuAspTyr 25
 Db 1704 GCAGCAGCTGCACGACGACGAGCGGACGACGACCTCCACGACATTTGCTT 1645

QY 26 GlnArgAla-----AlaAlaLeuGlyGlyProGluAspGluProGlyAlaAlaGluAla 43
 Db 1644 CTCAGAGCCCTTCTAGTCTCTTCTCTCCACGCTGACAGCGGCGGCTGCA 1585

QY 44 HisPheLeu-----ProArgHisArgLysLeuLys 53
 Db 1584 GCAGCTGGCGACATCCGCCACCAAGCCGCTAGTCTCCAGTCCACCAAAACCTTCT 1525

QY 54 GluProGlyProProLeuAlaSerSerGlnGlySerProAlaProSerProAlaGly 73
 Db 1524 CCACCTGCACGACGACGACGCTAGTGCAGCGGCTAAAGCTCCGCTCCGCTCCACCTCA 1465

QY 74 CysGlyGlyLysSerLysArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlu 93
 Db 1464 AGTCCACCTTCGGCTCTCTGCAGCAGCGAGACAGCTAGAGCTCCAGCTCCACCGCCG 1405

QY 94 GluSerTrpGlyLysSerValProLeuProCysProProProAlaThrLysGlnAlaGly 113
 Db 1404 CTCCTCTCTCCGTTACCACTCTCCGTTACCAACGCGCTCCACCTCAGCTGAAGCTGAT 1345

QY 114 IleGlyGlyLysLeuProAlaAlaAlaGlyLysSerProArgProLysTyr----- 131
 Db 1344 GCAGATCCAGAT---GCTGAAGCAGATGCTGAGCTTTTCAATGAGACTGCAGATCCTT 1288

QY 132 -----GlnAlaValLeuProIleGlnThrGlySerLeu 142
 Db 1287 AAGACATCTTCAATGCTCAATTTGAGATTTTGTCTCTACTGCTGATGCTTAGCA 1228

QY 143 ValAlaAlaAlaLysGluPro-----ThrProTrpAlaGly 154
 Db 1227 GAGCTTTTTCAGAGGCTCCATTCTCTGAGAGGCTTTAATGTCCTTAATGCGTCC 1168

QY 155 -----AspLysGlyGlyAlaAlaSerProAlaAlaThrAla 166
 Db 1167 CTTATACCACTATTGTCTAGAGCTGAGAGCAGATGACAGCAGCGGTGAGCGGACGC 1108

QY 167 SerAspProAlaGlyProProProLeuProLeuProGlyProProProAlaAlaProThr 186
 Db 1107 GCAGCAGCAGCGGCTCGCTCTCTCCACGCTCCACCTCCCTCCGACACGACGC 1048

QY 187 AlaThrAlaGlyThrLeuAlaAlaSerGlyLysArgTrpLysSerMetArgLysSerPro 206
 Db 1047 GCTGGCGGCGACCTGCAGCAGCGGCGACGACG-----GCACGT 1009

QY 207 LeuGlyGlyGlyLysSerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnIleLeu 226
 Db 1008 CTGCG-GCTCTCTCTCTCTGACCGGCTTACTCTCTGTCACCGTAGAAACCAACACC 950

QY 227 LeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnLeuGlnAlaLysGlnGlu 246
 Db 949 ACCGA-GTCTCCACCTCCGCGCAGCGGCGGACGACGACGACGCGGCGACGACATC 891

QY 247 IleGlnGluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArg 266
 Db 890 CTCATATCTCTCGAGGCGACCAAGAC-----CTCGAGGCTCCCAAGTCCGCCACCA 837

QY 267 ArgMetGlnLeuValLysLysAspAsnGlnLysGlnArgHisLysLeuPheGlnGlyTyr 286
 Db 836 GTCCTCCAGTCT-----CGCCACGAGTCTCCAAATGCCAC 801

QY 287 -----GluThrGlnGluArgGlnGluThrGlnLeuLeuSerGlu----- 298
 Db 800 CTGCGCTCTCTGACGCGGCGACGAGCGGCTGCTGACGACGACGCTCTCAAGTCTAC 741

QY 299 ---LysIleLysLeuGlyCysGlnProGluLeuSerGlnThrSerGlnThrLeuProPro 317
 Db 740 CTGCGCTCTCTCTGACGCGGCGACGAGCGGCTG---CGGCTGGGCGACGA-GCTCCACCA 685

QY 318 LysProPheSerCysGlyArgSerGlyLysGlnHisLysArgLysSerProPheGlySer 337
 Db 684 ACTCCACCCCGCGGCGGACGAGCGGCGAGTGCAGACACACACACACACACCT-----CCA 631

QY 338 ThrGlnArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLys 357
 Db 630 GCTCCACGACCTCCACGAGGAGGAGGCTGACCAAGATCATCG-----AAATGAAA 577

QY 358 ThrProLysHisSerProIleLysGlnGluProCysGlySerLeuSerGlnThrValCys 377
 Db 576 TCAATTAATGCTCTCCACCGCCCAAGTCCCA----- 544

QY 378 LysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrPro 397
 Db 543 -----AGACCTCCAAAGTCCGCAAGTCCCAAGTCCACCTCCGAGACCT 508

QY 398 ProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGlnLysAlaPheSer 417
 Db 507 CCGGCCAATCAAGTCCGAGTCCAGCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 448

QY 418 SerGlnIleGlnAspLeuProTyrLysSerThrThrGluMetTyrLeuCysArgTrpHis 437
 Db 447 GCACCA----- 442

QY 438 GlnProProProSerProLeuProLeuArgGluSerSerProLysLysGlnGluThrVal 457


```

Qy 95 SerTrpGlySerValProLeu-----ProCys-----ProProAlaThr 109
Db 337482 AAGCGGAGCGCGGAGCGCGGAGCGCGCTTGTGATGCGGCTCGCGGCGGACCGCG 337423
Qy 110 LysAlaAlaGlyIleGlyGlyGluProAlaAla---AlaGlyAla-----Gly 124
Db 337422 GCACGGCGCGGCTGGCTGGACACCGCGCGCGCGCGCGCGCGGCGGCGGCT 337363
Qy 125 CysSerProArgProLysTyrGlnAlaValLeuPro-----IleGlnThrGlySerLeu 142
Db 337362 TGTTCCTCGGTGGCGCGAGGAGCGCGCGCGCGCGCTTCCCAAAACTTATCGGTG 337303
Qy 143 ValAlaAlaAlaGlyGluPro-----ThrProTrpAlaGlyAspLys 156
Db 337302 CCGCGCGACCGCGCGGAGCGCGGTGTACCGGTGGCTGTTCGACAGCGCGCGCGCG 337243
Qy 157 GlyIlyAla-----AlaSerProAlaAlaThr-----AlaSerAspPro 169
Db 337242 GGGCGGCGGGGTTGGCGGCAATGACCGGGGAGACGGATGTGCTTCGCGCGCG 337183
Qy 170 AlaGlyPro-----ProProLeuProLeuProGly----- 179
Db 337182 GCGCGACCGCGCGGGGCGGACGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 337123
Qy 180 -----ProProLeuAla 184
Db 337122 TGTTCGCGCGCGCGGACACTGTGCGGTGGCGGACAGCGCGCGCGCGCGCGCGCA 337063
Qy 185 ProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGly---ArgTrpLysSerMetArgLys 204
Db 337062 ACGCGGCGACGCGGCGGACGCGCGCTTACTTCGCGCGCGCGCGCGCGCGCGCG 337003
Qy 204 s-----SerProLeuGlyGlyGlyGlyGly 213
Db 337002 GCACGGCGCGGAGCGCGCTTATGTCGCGCGGAGCGCGCGCGCGCGCGGAGAG 336943
Qy 213 rGlyAlaSer-SerGlnAlaAlaCysLeu-LysGlnIleLeuLeuLeuGlnLeuAspLeu 232
Db 336942 TGGCGGCTGCTTTCGCGCTTTCGCGGCGGCGCGGCGGCGCGCGCGCGCGCGCG 336889
Qy 233 IleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 252
Db 336888 ---CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 336832
Qy 253 GluArgAspThrLeuLeuAlaAlaArgIleGluArgMetGluArgMetGlnLeuValLys 272
Db 336831 CGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336772
Qy 273 LysAspAsnGluLysGluArgHisLysLeuPheGlnGlyArgGluArgGlu 292
Db 336771 CGCGACACGCGGTGCTGATCGGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 336712
Qy 293 GluThrGluLeuSerGluLysIleLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 312
Db 336711 CACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 336652
Qy 313 GlnThrLeu-----ProProLysProPheSerGlyArgSerGlyLysGly 328
Db 336651 GTACCGCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336595
Qy 329 HisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAla 348
Db 336594 ATCCCGACGATCAGCGCTGCGCTGCTTCACAGACACACCGCGGTTCACCGCGGTAC 336535
Qy 349 ProGluPheSerLysValLysThrLysThrProLysHisSerProIleLys----- 365
Db 336534 AAGTCGCTTACCAAGATTTC-ACCAAAACCCCGAGGTGTTCGCTCGCAATGGGATGC 336476
Qy 366 -----GluGluProCysGlySerLeuSerGluThrValLys 377
Db 336475 TTCTTACTGGCGATTTTTGTGCTGTGACGCTTCCCGCGGTGCGGATGACAGACGCTCATGC 336416

```

```

Qy 378 -----LysArgGluLeuArgSerGlnGluThr 386
Db 336415 GTATTGGAGAGTGTCAGATGTCGTTGTGATGGCGGCGGAGAGGTTATCGCGGACGCG 336356
Qy 387 -----ProGluLysProArgSerSerVal---AspThrProProArgLeuSer 401
Db 336355 CAAAGGATTGGCCAGTCGTGCGATGAGCATACCGCGCGGCAACGCGCGCGCGCGCA 336296
Qy 402 ThrProGln-----LysGlyProSerThrHisProLysGluAlaPheSer 417
Db 336295 ACACCGAGGACATATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336260
Qy 418 SerGluIleGluAspLeuProTyrLeuSerThrGluMetTyrLeuCysArgTrpHis 437
Db 336259 -----TGTGCACACCA 336248
Qy 438 GlnPro-----ProProSerProLeuProLeuArgGlnSerSerProLysGlu 454
Db 336247 TCGCGCGCGCTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336191
Qy 455 GluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAla 474
Db 336190 -----AGCGCTTTCATCCCAATTCGTCGACGCGCTTACCTCCGCTGCGCGCG 336143
Qy 475 ValProSerTrpArgAspHisSerValGluProLeuArgAspProAspPro-SerAspLe 494
Db 336142 -----CCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 336113
Qy 494 uLeuGlnAsnLeuAspSerSerValPheSerLysArgHisAlaLysLeuGluLeuAspG 514
Db 336112 CGCGCGCTGATCGATCAACAGATTCTTCGCGCAATACCGCGCGCGCGCGCGTGTGATCG 336053
Qy 514 uLysArgArgLysArg-----TrpAspIleGlnArgIleArgG 527
Db 336052 GCACCGCGCGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336008
Qy 527 uGlnArgIleLeuGlnArgLeuGlnLeuArgMetTyrLysLysLysGly 543
Db 336007 ACGCGGCTGTGATTATCGGCAACGCGCGCGCGCGGAGATCGCGCGCGCGCG 335959

RESULT 14
US-08-864-038A-1/c
; Sequence 1, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; OPERATING SYSTEM: IBM compatible
; SOFTWARE: Microsoft Windows 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389

```



```

Db      2511 AGGACTAAAGCCACCATGCTGAGCTGCGGCTGAGTGTGCGGACGACACTACCTATC 2570
Qy      404 -----GlnysglYProSerThrHisProLysglYValAlaPheSerGluIle 420
Db      2571 TGAGGCTGAATCGGGGCCGCTCTGCTCCACATCGTCGCC-----AGTGAACCG 2624
Qy      421 GluAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTrpHisGlnProPro 440
Db      2625 CAAATTCACACCTTCTTGCGGGACCATCTCGGCCACACTGTCAATGCGACCCGCTCTGC 2684
Qy      441 ProSerProLeuProLeuArgGluSerSerProLysglYValAlaValAlaArgCys 460
Db      2685 CCAAGAGAGTTCCAAAGAGAGGGGCCACACCGAGAGCTTCAG-----2726
Qy      461 LeuMetProSerSerValAlaIglYValThrSerValLeuAlaValProSerTrpArgAsp 480
Db      2727 -----TAAAGCGCCAGAGCTCAAGTGGACATGCTCCTCGGAGCGAGCGAT 2768
Qy      481 HisSerValGluProLeuArgAspProAsnProSerAspLeuGluAsnLeuAspAsp 500
Db      2769 GATTGACATCGTCATGCGAGAGATGAC-----2795
Qy      501 SerValPheSerTyrSargHisAlaLys-IleGluLeuAspGluLysArg-----Argly 518
Db      2796 -----CATGTCAACCTGAGAGCGCTGACATGAGCGGCTCATCAAGAA 2837
Qy      518 SarGTrpAspIleGluArgIleArgGlu-----GlnArgIleLeuGluArgLeuGlnLe 536
Db      2838 AAGGGAGAGCTTCTCTCTGCGAGGACACTGCGGAGAGCGGAGAGCGGCTGCGGC 2897
Qy      536 uArgMet---TyrLysLysLysglYIleGlnGluSerGluProGluValThrSerPheph 555
Db      2898 TGAGAGCCCCGAGAGAGAGAGGGGCTGCGAGCGCTGCGAGAGATCGAGTGCCTGCGC 2957
Qy      555 eProGluProAspAspValGlu-SerLeuMetIleThrProPheLeuProValAlaLap 575
Db      2958 AGCCAACTGATGACTACATCATGATGACGATCACCGCTGCCAGGCCACATCGTCAGCT 3017
Qy      575 heGlyArgProLeuProLysLeuThrProGln-----AsnPhGluLeuP 590
Db      3018 GGGAGAGACCCAGAGAGAGCTGAGCTCCACAGACACATCCGTGTCATCGCTCGCTC 3077
Qy      590 roTrPLeuAspGluArgSer 596
Db      3078 CTTGGCTGAAGCCCGGCTCC 3097

RESULT 12
US-09-718-815-1
; Sequence 1, Application US/09718815
; Patent No. 6455293
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6455293el motor proteins and methods for
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718, 815
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4911
; TYPE: DNA
; ORGANISM: Human
US-09-718-815-1

Alignment Scores:
Pred. No.: 0.0216 Length: 4911
Score: 212.50 Matches: 181
Percent Similarity: 34.05% Conservative: 88
Best Local Similarity: 22.91% Mismatches: 261
Query Match: 6.56% Indels: 260

```

```

DB:      4      Gaps:      36
US-10-054-935-2 (1-614) x US-09-718-815-1 (1-4911)
Qy      14 ProAlaGly---GlyAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
Db      897 CCTGCTGGCCCTTGCGCAATGTCATGACGCG-----CTTAGG 932
Qy      33 GlyProGluAspGluProGlyAla-----AlaGluAlaHisPheLeuPro 47
Db      933 GGACCAAGACCAAGAGGTGTGACCTTCCCTTACAGGACTCCAGACTCTAGCTGCTCCT 992
Qy      48 -----ArgHisArgLysLeuLysGluPro-----55
Db      993 CCAGATTTGCTGGGGGCAACAGCCAGACCATCATGATGATGCTGTGTGAGCCCTCAG 1052
Qy      56 -----GlyProProLeuAlaSerSerGlnGlyLysProAlaProSerPro 71
Db      1053 CCGAGATTTCATGAGAGACCT---CAACACACTCAATATATGCCAATCGGCCCGCAACAT 1109
Qy      72 AlaGlyCysglYglYLysglYArgglYLeuLeuProAla-----85
Db      1110 CAAGAACAGAGTGTGTAGTACACGACGACCAAGACCAGCCAAATCATGCTACTGGGGC 1169
Qy      86 -----GlyAlaAlaProglYglYglYglYglYglYglYglYglYglYglYglY 95
Db      1170 TGAGATTGCTCGGCTGAGATGAGATGAGATGATGATGATGATGATGATGATGATGATG 1229
Qy      96 TrpGlyGly-----SerValProLeuProCysProPro-----106
Db      1230 AGAGGATGGCGCTGAGGGCTATAGTATGATGATGATGATGATGATGATGATGATGATG 1289
Qy      107 -----ProAlaThrLysGlnAlaGlyIleGlyGlyGluProAlaAlaAlaGlyAla 123
Db      1290 GAATGGGGCCCTCGGCT-----GCGGGTGAAGGCCATGCGAGAGGCCAT 1334
Qy      124 GlyCysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGlySerLeuVal 143
Db      1335 CGATGCCATCAACACACCGGCTTACCCACGCTCATGAGGCA-----GSAGGCCAACCT 1385
Qy      144 AlaAlaAlaLys-----GluProThrProTrp 152
Db      1386 GCTGTAGTACCAAGCCGCGGATGCAATGAGGCCATTGTGCTGCTGATTCACAGACTACAT 1445
Qy      153 AlaGlyAspLysGlyAlaAlaAlaSerProAlaAlaThrAlaSerSerProAlaGlyPro 172
Db      1446 CCGGGAGATGAGAGAGCTACGACTTAAGCTTTAGAGAGTGAAGCCATGAACGATGCCCT 1505
Qy      173 ProProLeuProLeuProGly-----179
Db      1506 GCGCCGAGCTCTCAAGGGCTCGGCTAGAGCCCTTACTCCTGCGTCTTCACAGC 1565
Qy      180 ProProProLeuAla-----Pro 185
Db      1566 CGCCCGGCTTGGGGGAGCCCTGCGACCTTCATGAGAGATCCTCGAGGTATCCG 1625
Qy      186 ThrAlaThrAlaGlyThrLeuAlaAlaSer-----195
Db      1626 CAGGCCACAGCAGAGCTGAGCGGCTAAAGAAAGAGAGGTACAGCAGCGAGGAGAGAG 1685
Qy      196 -----GluGlyArgTrpLysSerMetArgLysSerProLeuGly-----208
Db      1686 CCCCAGAGAGAGAGCTTCAAAAAGAGGCAAAATCCAAACAGAGAGAGAGAGAGAGAC 1745
Qy      209 -----GlyGlyGlyGlySerGlyAlaSer-SerGlnAlaAlaCysLeuLys---223
Db      1746 GGATGAGAACAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Qy      224 -----GlnIleLeuLeuGlnLeuLys 231
Db      1806 GGAAGGGCGCGAGATGAAGATGAGAGACTCGGCACTGAAGAGAGAGCTGTGATCAGCA 1865
Qy      231 pleuIleGluGlnGlnGln-----GlnIleLeuLeuGlnLeuLys 239

```


QY 56 -----glyProLeuAlaSerSerGlnGlySerProAlaProSerPro 71
 Db 1053 CCGAGATTTCATGGAGACCT---CAACACACTCAAAATATGCCAATGGCGCCCGCACAT 1109
 QY 72 AlaGlySerGlyGlyLysGlyArgGlyLeuLeuProAla----- 85
 Db 1110 CAAAGACAAGGTGGTAGTGAACCAAGACCAAGCAAAATCAATGATCGCGGC 1169
 QY 86 -----GlyAlaAlaProGlnGlnGlnGlnSer 95
 Db 1170 TGACATTGCTGGCTGCCAGATGAGCTGATGAGCTAATAGCGGCGCAAGCATGATAGG 1229
 QY 96 TrpGlyGly-----SerValProLeuProCysProPro----- 106
 Db 1230 AGAGATAGCGCGCTGAGGCGCTAGTGAATCTCTCCGAGAGAAATGCCATGCTACAGAAAGA 1289
 QY 107 -----ProAlaThrLysGlnAlaGlyTleGlyGlyLysProAlaAlaAlaGlyAla 123
 Db 1290 GAATGGGCGCTCGCGCT-----GCGGCTGAAGACCATGACGAGAGCCAT 1334
 QY 124 GlySerSerProAlaGlyProLysThrGlnAlaValLeuProGlnThrGlySerLeuVal 143
 Db 1335 CGATGCCATCAACAACCGCTCACCACCTCATAGCCA-----GGAGGCCAACCT 1385
 QY 144 AlaAlaAlaLys-----GluProThrProTrp 152
 Db 1386 GCTGCTAGCCCAAGCGCGCATGGAATGAGCCATTGGTGGTGATGCCAAGCATCAT 1445
 QY 153 AlaGlyAspLysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyPro 172
 Db 1446 CCGGAGATTCAGAGGACTACGACTAAGCTTCTAGAGAGTAGAGCATGAACGATGCCCT 1505
 QY 173 ProProLeuProLeuProGly----- 179
 Db 1506 GCGCGCGAGCTCTCAGCGGCGCTCGGCTAGAGAGCCCTACTCCTGGTCTTCCAGC 1565
 QY 180 ProProLeuAla-----Pro 185
 Db 1566 CGCCCGCGCTTCGCGGGGCAACCTGCCAGCTCATGAGAGATGCGCTCGAGGTGATCCG 1625
 QY 186 ThrAlaThrAlaGlyThrLeuAlaAlaSer----- 195
 Db 1626 CAGGCCCAAGCAGACCTCGAGCGGCTAAAGAGAAGAGAGGCTCAGCGCAGAGAACG 1685
 QY 196 -----GluGlyArgTrpLysSerMetArgLysSerProLeuGly----- 208
 Db 1686 CCGGAGAGAGAACCTTCAAAAAGAGGCAAACTCCACAGAGAACACCGAGAGAC 1745
 QY 209 -----GlyGlyGlyGlySerGlyAlaSer-SerGlnAlaAlaCysLeuLys-- 223
 Db 1746 GGATGAGAACGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGTGGCTGTGAGAGGA 1805
 QY 224 -----GlnIleLeuLeuGlnLeuAs 231
 Db 1806 GGAAGGCGCGAGATGAGATGAGACTCGGCGCATGTAAGAGAGCTGTGGTGCATCAGA 1865
 QY 231 PheIleGlnGlnGlnGln-----Gln 239
 Db 1866 CTCAGACCCCGAGAGAGAGAGGCTGCACTCCAGCGGACCTGGCCGACCTGATTGGA 1925
 QY 239 nLeuGlnAlaLysGlnLysGlnIleGlnGlnLeuLysSerGlnArgSerThrLeuAla 259
 Db 1926 GATCGAATCAAGAGAGAGCTGATGAGAGCTGAGAACAGCAGCGCGCGCTTGCAAGC 1985
 QY 259 aArgIleGlnArgMetGlnArgArgMetGlnLeuValLys-----LysAspAsnGln 276
 Db 1986 GCTCAACGACGAGATAGGAAAGCTGATTCCTGACAGAAACAAGATCCAGACACACA 2045
 QY 276 uLysGlnArgHisLysLeuPheGlnGlyTyrGlnThrGlnGlnArgGlnGlnGln 296
 Db 2046 GCTGAGCGGACCGCTGTCTGTCGCAAGAACCTCAGACCATGAGTGTCTATCTGAGAGAA 2105

QY 296 uSerGlnLysIleLysLeuGlnCysGlnProGlnLeuSerGlnThrSerGlnThrLeuPr 316
 Db 2106 GGCCAAAGATTCAGAGGAGACTATGAGAAAGAGCTCGGAGATGAAACCGGACCTGCA 2165
 QY 316 oProLysProPheSerGlyArgSerGlyLysGlyHisLysArgLysSerProPheGln 336
 Db 2166 GAACCTCAGCGCCCGCAAGAAAGACAGCGCCGCTCTTAAGAACAGTCCGCGCTAC-- 2223
 QY 336 ySerThrGlnArgLysThrProValLysLysLeuAlaProGlnPheSerLysVal----Ly 355
 Db 2224 -----GAGAGGAG-----CTGAAGAACTCAGAGCGCGAGTGGCTGATGAGAA 2270
 QY 355 sThrLysThrProLysHisSerProIleLysGlnGlu---ProCysGlySerLeuSerGln 374
 Db 2271 GGCCAAAGTGGCCCTGATGAGACAGATGCTGAGAGACCAACAGCGGCGCGCTAGTGA 2330
 QY 374 uThrValLysLysArgGlnLeu----- 381
 Db 2331 GACCAAGAGAAACGGGAGATCGCACAGCTCAAGAAAGAGCAGCGCGCAGAGTTTCA 2390
 QY 382 -----ArgSerGlnGlu-----ThrPr 387
 Db 2391 GATCCGAGCTGTGAGTCCAGAAAGCGGACAGAGAGATGCTCTGAGAGAAAGACCA 2450
 QY 387 oGlu-----LysProArgSerSerValAspThr 396
 Db 2451 GGAGGTTTCTGACTGAGCGCGCTCGGCAAGCCATCTGAGCGGGTGGAGCGCGCTGC 2510
 QY 397 -----ProProArgLeu-----SerThr-Pro----- 403
 Db 2511 AGGACTAAAGCCACCATGCTGAGCTGGGCTGAGGTGCGGCCAGCATCTACTATC 2570
 QY 404 -----GlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSer-GluIle 420
 Db 2571 TGAGGCTGAATCAGAGGCGCGCTGTCTCCAGATGTGCGC-----AGTGAACCG 2624
 QY 421 GluAspLeuProLysLeuSerThrThrGlnuMetLysCysArgTrpHisGlnProPro 440
 Db 2625 CAAATCAACACTTCTTGGGGGACCATCTGCGCCACCATGTCATGAGCAACCGCTCTGC 2684
 QY 441 ProSerProLeuProLeuArgLysSerSerProLysLysGlnGlnThrValAlaArgCys 460
 Db 2685 CCGAAGAAATTCAGAGAAAGGCGCGCAGCAAGAGCTTCAG----- 2726
 QY 461 LeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValProSerTrpArgAsp 480
 Db 2727 -----TAAGCGCGCAAGCTCAAGTGGCAGTCCCTGGAGCGACGAT 2768
 QY 481 HisSerValGluProLeuArgAspProAsnProSerAspLeuGlnAsnLeuAspAsp 500
 Db 2769 CATTCAGATCGTCAATGCAAGAAATGAC----- 2795
 QY 501 SerValPheSerLysArgHisAlaLys-LeuGlnLeuAspGlnLysArg-----Argly 518
 Db 2796 -----CATTCCAACCTGAGAGCTGAGATGAGAGCGGCTCATCAAGAA 2837
 QY 518 sArgTrpAspIleGlnArgIleArgGlu-----GlnArgIleLeuGlnArgLeuGln 536
 Db 2838 AAGGAGAGAGCTGTCTCTCTCGAGAGAGGACCTCGAGAGAGGAGGAGGCGCTGCAAGC 2897
 QY 536 uArgMet---TyrLysLysLysGlyIleGlnGlnLysSerGlnProGlnuValThrSerPhe 555
 Db 2898 TGAGAGCGCCGAGAGAGAAAGAGGCGCTGAGAGAGCTGAGAGAGATGCAAGGTGCTGGC 2957
 QY 555 eProGlnProAspAspValGlu-SerLeuMetIleThrProPheLeuProValAlaAla 575
 Db 2958 AGCCAACTTGATCAATCAATGACGCGCATCACCAGCTGCCAGGCCACTGTCGACGT 3017
 QY 575 heGlyArgProLeuProLysLeuThrProGln-----AsnPheGlnLeu 590
 Db 3018 GAGAGAGACCAAGAGAGAGCTGACTCCACAGACATCCGTGTGTCATGCTCTCTC 3077
 QY 590 roTrpLeuAspGlnArgSer 596

OY	89	ProGlyInIleInIleuInIleuSerTPrgIyGlySerValProLeuProCysProPro-----	106
Db	6199	CTGGGACCACTGGTAAGCCCTGGCGGGGCGAGAGCCCTTCGGGGCTCTCTGACCAT	6258
OY	107	-----ProIleThrLysGlnAlaGlyIleGlyGlyGluProAlaAlaIleAlaIleGlyAlaGly	124
Db	6259	GGTGGCCCGCCG---GCCAGCGCGGAGACGGGAGAGAGGCCCGCCCTCTTGGGGCGGA	6315
OY	125	CysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAla	144
Db	6315	-----	6315
OY	145	AlaAlaLysGluProThrProThrProAlaGlnAspLysGlyGlyAlaAlaSerProAlaAla	164
Db	6316	---GGGCGTTCCCTCTTACAGAGCGGTGTGAGCGGGGCGG-----	6354
OY	165	ThrAlaSerAspProAlaGlyProProPro-----LeuProLeuProGlyProProPro	182
Db	6355	-----AGCGTTCCCAAGCGCGCTCTCTGGGGGCGCTGACCTCAACGGGTGCCGAGCC	6408
OY	183	LeuAlaProThrAlaThrAlaThrAlaThrLeuAlaAlaSerGluGlyArgTrrLysSerMet	202
Db	6409	CAGGC---GAGCGTGGCTGGGGGAGACTACCGGGGTGGGGAGA-----	6453
OY	203	ArgLysSerProLeuGlyGlyIleGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeu	222
Db	6454	-----GCCTCGTAAGAGGTGG-----GGGATGACCAAGTTCCGACGCTCTCT	6498
OY	223	LysGlnIleLeuLeuLeuGlnIleuAspLeuIleGluGlnGlnGlnIleuGlnAla	242
Db	6499	AAAGAGGT-----	6507
OY	243	LysGluLysGluIleGluGluLeuLysSerGluThrGlyAspThrLeuLeuAlaArgIleGlu	262
Db	6508	-----AAGGAGAAGAGAGAGGCCCTCCGGGAGCGGCTCGGGGGAAGAGCGGCGGAG	6561
OY	263	ArgMetGluThrArgArgMetGlnLeuValLysLysAspAsnGluLysGluArgHisLysLeu	282
Db	6562	GACTACGTGGCCATGAGAGCTACTACAGCAA-----GAGTTCAACGGAGGCTGAG	6615
OY	283	PheGlnIleTyrGluThrGluGluArgGluLeuThrGluLeuSerGlyLysIleLysLeu	302
Db	6616	CTGAAGGCCCTGGAG---GAGGAGAAGAGCTTTCGGAGCGGTGGAGAGAGGCCCTGAG	6672
OY	303	Glu-CysGlnProGluLeuSerGluThrSerGlnThrLeuProProLysProPheSerCys	322
Db	6673	AAAGTCTGGGTGACTCTC-----CTCCGCGCTCGT-----	6703
OY	322	sgLYArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrAlaGlyThr	342
Db	6704	-----AGAGCGTGAAGAAGAGGTAAAGAGATC-----GTCAACCTTAACAATC	6747
OY	342	rProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHis	362
Db	6748	GCCCTGAAGACCTC-----GAGCTTCAACGGAGGCTGAG	6777
OY	362	rProIleLysIleuProCysGlySerLeuSerGluThrValCysLysArgGluLeuArg	382
Db	6778	CCCC-----CAGAGGCGGCGCTCTTTCGAAGCGCTGATCCGAGATATGA	6828
OY	382	gSerGlnGluThrProGluLysProArgSerSer-----ValAspThr	396
Db	6829	AGG-AACTCCACCCCATCTGGAGCGAAGATGTGGTATGCTCCCTTGTCGAGCAAC	6887
OY	396	rProPro---ArgLeuSerThrProGln-----LysGlyLys	407
Db	6888	GCCACCGGCAAGGGGCTCAGGGCAGCGAGATCGGCGTGGGCTGAGGCTACGGAGAG	6947
OY	407	oSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeu	427
Db	6948	GAGCAAAAGGTATACGCGATCCAGGTCAAGCTGTGGATTAAG-----CCCTCTTCTGG	7001

[illegible]

RESULT 10
US-09-718-692-1

```

: Sequence 1, Application US/09718692
: Patent No. 6383796
:
: GENERAL INFORMATION:
:
: APPLICANT: Beraud, Christophe
: APPLICANT: Freedman, Richard
: TITLE OF INVENTION: No. 6383796e1 motor proteins and methods for
: TITLE OF INVENTION: their use
: FILE REFERENCE: 1052
: CURRENT APPLICATION NUMBER: US/09/718,692
: CURRENT FILING DATE: 2000-11-22
:
: NUMBER OF SEQ ID NOS: 4
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 1
:
: LENGTH: 4911
:
: TYPE: DNA
:
: ORGANISM: Human
:
: US-09-718-692-1

```

Alignment Scores:

Pred. No.:	0.0216	length:	491
Score:	21.50	Matches:	18
Percent Similarity:	34.05%	Conservative:	18
Best Local Similarity:	22.91%	Mismatches:	26
Query Match:	6.56%	Indels:	26
DB:	4	Gaps:	36

US-10-054-935-2 (1-614) x US-09-718-692-1 (1-4911)

QY	1	PROIAlAGLY---	GLYASNP	ROGLI	ALARG	LEUASNP	SPYR	GLU	IRG	ALAL	ALAL	LEU	GLY	32					
Db	897	CTGCGTGGCGCT	GGGCGCAT	GTGATC	AGCAGGC	-----	-----	-----	-----	-----	-----	-----	CTTAG	932					
QY	33	GLYP	ROGLI	ASNP	GLI	PL	ROGLI	LYALA	-----	AL	GLU	AL	ASP	LEU	PRO	47			
Db	933	GGACCA	GAGCA	GAA	GAAG	GTGTG	TCAC	CGTTC	CCCTAC	AGGAG	ACTCC	CAAC	ACTCT	AC	CGGCTCT	992			
QY	48	-----	-----	-----	-----	-----	-----	-----	-----	Arg	His	Arg	Lys	Leu	Lys	GLU	Pro	-----	55
Db	993	CCAGGATTC	GCTGG	GGGG	CAAC	AGCC	AGCAC	AGAC	CAATC	ATG	ATG	CGCTG	TGTG	TGAGG	CCCTC	GAG	1052		


```

1 APPLICATION NUMBER: US 60/009,704
2 FILING DATE: 11-JAN-1996
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 60/001,995
5 FILING DATE: 01-AUG-1995
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Chao, Mark
8 REGISTRATION NUMBER: 37,293
9 REFERENCE/DOCKET NUMBER: 95,963-E
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: 312-913-0001
12 TELEFAX: 312-913-0002
13 INFORMATION FOR SEQ ID NO: 7:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 3147 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: cDNA
20 FEATURE:
21 NAME/KEY: CDS
22 LOCATION: 245..1231
23 OTHER INFORMATION: /note= "Tsps E101 sequence longest
24 open reading frame; other possible start codons are TTG/Leu
25 OTHER INFORMATION: TTG/Leu13; TTG/Leu15; GTG/Val143"
26 FEATURE:
27 NAME/KEY: mat_peptide
28 LOCATION: 245..1231
29 US-08-781-802-7
30
31 Alignment Scores:
32 Pred. No.: 0.00939 Length: 3147
33 Score: 215.50 Matches: 157
34 Percent Similarity: 30.04% Conservative: 53
35 Best Local Similarity: 22.46% Mismatches: 193
36 Query Match: 6.66% Indels: 298
37 DB: 2 Gaps: 29
38
39 US-10-054-935-2 (1-614) x US-08-781-802-7 (1-3147)
40
41 QY 4 ArgSerAlaValAlaPheLysAlaAlaAlaPro-----AlaGly 16
42 Db 2643 AAGAGCCCTTAAGAGAGAGCGCCCTCGCCCGCCCTCCCGCCGCGCCGCGCCAGG 2584
43 QY 17 GlyAsnProGluGlnArg-----LeuAspTyrGluArgAlaAlaLeu 31
44 Db 2583 GCCCTCCCTCGAAGAGCGCGCCGCTCCCAAGCCGCGCCAGAGGAGGCGCTCTGG 2524
45 QY 32 GlyGlyProGlu-----AspGluPro 38
46 Db 2523 GGAAGAGCCCGAGATGATCGCCGCGCCTGGAGAGCCCGCGGCGCTCTCGCCCTTCCC 2464
47 QY 39 GlyAlaAlaGluAlaHisPheLeu-----ProArgHisArgLysLeu 52
48 Db 2463 GCCCTCCAGGGCGACCCACCCCTTGAGCGCGCTCCACCTCGCGCCCGCGCCAGGCGG 2404
49 QY 53 LysGluProGlyProProLeuAlaLaserSerGlnGly---GlySerProAlaProSerP 71
50 Db 2403 GCCCTCCAGGGCGCTCCCGCGCGCCTCGCGCGCCGCGCCAGAGAGAGAGCGCGCCG 2344
51 QY 71 AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyG 91
52 Db 2343 TTCAGGGCGACGACCTCCGGCGGGGGGCGCTCCCGCTCCAGCGGGCGCAGGCGCTCTCG 2284
53 QY 91 ngIngluGluSerTTrpGlyGlySerValProLeuProCysProProProAlaThrLysG 111
54 Db 2283 TACTGCGCTCGCGCAGAGGAGGACTGCGGACGTCCAGCGCCCTGGGCGCAGG--- 2232
55 QY 111 nAlaGlyLeuGlyGlyGluProAlaAla-----Ala-G 122
56 Db 2231 -CCAGGCGCAGAGAGAGACACCCCGACGACGATACCTTCATGATACCGCTTTCGCGCG 2173
57 QY 122 LysAlaGlyCysSerProArg-ProLysTyrGlnAlaValLeuProIleGlnThrGlySer 141

```

```

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-19

Alignment Scores:
Pred. No.: 0.121 Length: 38506
Score: 217.00 Matches: 177
Percent Similarity: 31.13% Conservative: 63
Best Local Similarity: 22.96% Mismatches: 256
Query Match: 6.70% Indels: 282
DB: 3 Gaps: 36

US-10-054-935-2 (1-614) x US-09-320-878-19 (1-38506)
QY 1 MetThrMetArgSerAlaValPheLysAlaAlaAlaProAlaGlyAsnProGlu 20
Db 27161 GTCCACCTGGCGCAAGTCCATCGCCGCCCGCGCAAGCCCATGATCTCC 27220
QY 21 GlnArgLeuAspTyrGlnAlaGlnAlaAlaLeuGlyGlyProGlnAspGlnProGly 39
Db 27221 CTCGCCCTCGAC-----GAGCGCGCGCTCTGAAGCACTGAGCACTTCAGCGCACTC 27274
QY 40 ---AlaAlaGlnAlaHisPheLeuProArgHisArgLysLeuLysGlnProGlyPro 57
Db 27275 TCCGTGCGCCGCGCTCAACGG-CCCCACCGCCACCGCTGTCTCCGGCGCAACCCGACGAT 27333
QY 58 -----CysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGly 65
Db 27334 CGAAGAACTCGCCCGCACGCGGAGCGGCGGCTCGCGCGGATATCCCGGTCGA 27393
QY 66 SerProAlaProSerProAlaGly-----ProLeuAlaSerSerGlnGly 73
Db 27394 CTACGCTCCACAGCGCGGAGGTGCGAGATCATCGAAGAGAGCTGCGCGAGTCTCGC 27453
QY 74 -----CysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGly 90
Db 27454 CGGACTCGCCCGCGAGGCTCGCGACGTGCGCTCTCT-CCACCTCGAAGGACGTCGA 27512
QY 91 GlnGlnGlnGlu-----SerTrpGly 97
Db 27513 TCACCGAGCGGCGTCTGACGGCACTACTGTAACCGCAACCTGCGCATCGCGTGGCT 27572
QY 98 GlySerValPro-----LeuProCysProProProAlaThrLysGlnAla 112
Db 27573 TCCGCCCGCGCTGAGAGACCTTGCGGCTTGACGGCTTCACCCACTTCATGAGGTCAGCG 27632
QY 113 GlyIleGlyGlyLysProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTrpGln 132
Db 27633 -----CCACCCCGCTCTCAACATGACCTTCGCCGAGACCGTCA----- 27671
QY 133 AlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGlu----- 148
Db 27672 -----CGGGCTCGGCAACCTCC-----GCAGCAACAGGAGGCGCAGAGAGCTC 27716
QY 149 -----ProThrProTrpAlaGly---AspLysGlyGlyAlaAla 160
Db 27717 TGGTCACTTCATCGCGAAGCCTTGCGGCAACGCGCTCACCATGACTGCGGCCCATCC 27776
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProPro----- 173
Db 27777 TCCCAACCGCAACCGCGGACCGAGCTCCCACTACGCTTCACGAGACCGAGCGCT 27836

```

```

QY 174 -----ProLeuProLeuProGlyProProProLeuAlaProThrAlaThr 188
Db 27837 TCTGGCTGCAGACCTCCGCCCA-----CCAGCGCCCGCGACGACTGGCTTACCGGCTCG 27893
QY 189 AlaGlyThrLeuAlaAlaSerGlnGlyArgTrpLysSerMetArgLysSerProLeuGly 208
Db 27894 AGTGA-----ACCGCTGACGGCTCCGCCAGCGGACCTGTCCGCGCGGT 27941
QY 209 GlyGlyGlyLysSerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnIleLeuLeu 228
Db 27942 GGATCTCGCGCTCGGAGCGGACGACAGACCGAGCTCG----- 27982
QY 229 GlnLeuAspLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 248
Db 27983 -----GCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28015
QY 249 GlnLeuLysSer-----GluArgAspThrLeuLeuAlaArgIleGluArgMet 264
Db 28016 GTACTGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28074
QY 265 GluArgArgMetGlnLeuValLysLysAspAsnGlnLysGlnArgHisLysLeuPheGln 284
Db 28075 GACGACCGCGGAGCGCTTCACCGCGCGGTCTCTCTCTGACGAGCACTCTGCCACAGT 28134
QY 285 GlyTyrGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 304
Db 28135 CGGCTGGGTGCAGGCACTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28194
QY 305 GlnProGlnLeuSerGlnThrSer----- 312
Db 28195 GGGCGCGGTCTCTCGCGAGCTCTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28254
QY 313 -----GlnThrLeu-----Pro 316
Db 28255 GGGCTCGCGCGCGGTCTCGCTCTGACACCGCGAGCGTGGCGCGCGCTCTCGACCT 28314
QY 317 ProLysProPheSerCys-----GlyArgSerGlyGlyHis 329
Db 28315 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28374
QY 330 LysArg-----LysSerProPheGlySerThrGluArgGlyThrProValLysLys 346
Db 28375 CGCGGAGGACGATCGCATCTCCGACACCGCGGACTCTCCCGCGCGCGCGCGCGCGCGCG 28434
QY 347 LeuAlaProGlnPheSer---LysValLysThrLysThrProLysHisSerProIleLys 365
Db 28435 ACCCTTCACGAGAGTGGCGCCACCGCGAGCTGGACCGCCACGCGGACGCTCT- 28488
QY 366 GlnGlnProCysGlySerLeuSerGlnThrValLysLysArgGlnLeuArgSerGlnGln 385
Db 28489 -----CATCACCGCGCGCACCGG 28506
QY 386 ThrProGlnLysProArgSerSerValAspThrProProArgLeuSerThrPro----- 403
Db 28507 AGCCCTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28566
QY 404 -----GlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSer 418
Db 28567 CGTCAGCGGAGCGCGGAGAACACCGCGGAGGACACCACTCAACCG- 28614
QY 419 GluIleGlnAspLeuProTyrLeuSerThrGluMetThrLeuCysArgTrpHisGln 438
Db 28615 -----CGAAGTACCGGATCGGGCGCGCGCGTCAACATCG- 28649
QY 439 ProPro-----ProSerProLeu 444
Db 28650 CCGCTTCGAGAGTCCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28706
QY 445 ProLeuArgLysSerSerProLysLysGlnGlnThrValAlaArgCysLeuMetProSer 464
Db 28707 CCGAGAGCGCGCTTACCGCGCGGTCTTCACACCGCGCGCGCGCGCGCGCGCGCGCGCG 28765

```

```

QY 98 GlySerValPro-----LeuProCysProProAlaThrLysGlnAla 112
Db 29431 TCGCCCCCGCGGTGAGACCTTGGCGGCTTACACCTCATCATGAGGTCAGCG 29490
QY 113 GlyIleGlyGlyGluProAlaAlaAlaGlyLysSerProAlaProLysTyrGln 132
Db 29491 -----CCACCCCGCTCTCAACATGAGACCTCCCGAGACCGTCA----- 29529
QY 133 AlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGlu----- 148
Db 29530 -----CCGCGCTCGGACCCCTCC-----GCCGCGAAGAGGAGGCGCAGAGCGTC 29574
QY 149 -----ProThrProThrAlaGly-----AspLysGlyGlyAlaAla 160
Db 29575 TGGTCACTCACTCGCGGAGCGCTGGCGCCACAGCGCTCACTCACTGAGCGCCATCC 29634
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProPro----- 173
Db 29635 TCCCCACCGCAGACCGGCGCACCGAGCTCCCGACCTACAGCTTCAGACGAGCGCT 29694
QY 174 -----ProLeuProLeuProGlyProProProLeuAlaProThrAlaThr 188
Db 29695 TCTGGCTGAGAGCTCGCGGCCA-----CCAGCGCGCGCGAGCACTGGCGCTTACCGCTCG 29751
QY 189 AlaGlyThrLeuAlaAlaSerGlnGlyArgTyrLysSerMetArgLysSerProLeuGly 208
Db 29752 AGTGA-----AGCGCTGACGGCTCGCGGCGCAGGCGGACCTGTCGGCGCT 29799
QY 209 GlyGlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnIleLeuLeuLeu 228
Db 29800 GGAATCTCGCCCTCGGAGCGAGCGACAGACCGAGCTGCTG----- 29840
QY 229 GlnLeuAspLeuIleGlnGlnGlnGlnGlnGlnAlaLysGlnLysGlnIleGln 248
Db 29841 -----GGCGCGCTGAGAGCGCGCGGAGCGGAGCGGAGCTGAC 29873
QY 249 GlnLeuLysSer-----GlnArgAspThrLeuAlaAlaArgIleGlnArgMet 264
Db 29874 GTACTGGAAGCGCGGCGCGGAGAGCGAGCGCTGAGCGCTCGCGCGCTGAC--CGCACT 29932
QY 265 GlnArgArgMetGlnLeuValLysLysAspAsnGlnLysGlnArgHisLysLeuPheGln 284
Db 29933 GACGACCGCGCGGCTTCAACCGGCTGCTCGCTCTGACACCTCGGCGCAGAGT 29992
QY 285 GlyTyrGlnThrGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 304
Db 29993 CGCGTGGGTGCAAGCACTCGGCGAGCGCGGAGATCAAGCGCGCTGTCGTCGTCACCA 30052
QY 305 GlnProGlnLeuSerGlnThrSer-----GlnThrLeu-----Pro 316
Db 30053 GGGCGGCTCTCGTGGAGGCTCTCGACACACCGCGCGGAGCGCGGCGCATGCTCTG 30112
QY 313 -----GlnThrLeu-----Pro 316
Db 30113 GGGCGTGGCGCGCTGCTCGCTGAGACACCGCGGAGCGCGGCGCTGTCGACCT 30172
QY 317 ProLysProPheSerCys-----GlyArgSerGlyLysGlnHis 329
Db 30173 CCGCGCCACCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30232
QY 330 LysArg-----LysSerProPheGlySerThrGlnArgLysThrProValLysLys 346
Db 30233 CGGCGAGAGACAGATGCGCATCGCGACACCGGAGATCAAGCGCGCGCGCTGCGCGCGCG 30292
QY 347 LeuAlaProGlnPheSer-----LysValLysThrLysThrProLysHisSerProIleLys 365
Db 30293 ACCCTCCAGAGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30346
QY 366 GlnGlnProCysGlySerLeuSerGlnThrValCysLysArgGlnLeuArgSerGlnGln 385
Db 30347 -----CATCCGCGCGGCGCAGCGG 30364
QY 386 ThrProGlnLysProArgSerSerValAspThrProProArgLeuSerThrPro----- 403

```

```

Db 30365 AGCCCTCGGACCGACCGCGCGCGCGCGCGCTGTGAGTGGCCACACGAGCGGAACACCTCTCT 30424
QY 404 -----GlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSer 418
Db 30425 CCGTACCGCGAGCGGAGAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30472
QY 419 GlnIleGlnAspLeuProTyrLeuSerThrThrGlnMetTyrLeuCysArgTyrPheGln 438
Db 30473 -----CGAATCAACCGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30507
QY 439 ProPro-----ProSerProLeu 444
Db 30508 CGCGCTGGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30564
QY 445 ProLeuArgGlnSerSerProLysLysGlnGlnThrValAlaArgCysLeuMetProSer 464
Db 30565 CGGAGAGCGCGCTCAACCGCGCTCGTCCACACCGCGCGCGCGCGCGCGCGCGCGCGCG 30623
QY 465 SerValAlaGlyGlnThrSerValLeuAlaValProSerThrPargAspHisSerValGln 484
Db 30624 GACTCAACCGCGCGCGGAGACATCGCGCGATCTGGCGCGCGGAGAGAGCGCGCGCGCG 30683
QY 485 ProLeuArgAspProAsnProSerAspLeuLeuGlnAsn-----LeuAsp----- 499
Db 30684 GTCCCT-----GACGACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30728
QY 500 -----AspSerValPheSerLysArg 506
Db 30729 CTACTCTCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30788
QY 507 HisAlaLysLeuGln-----LeuAspGlnLysArgLysArg----- 519
Db 30789 AACGCCACCTCAACGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30848
QY 520 -----TyrAspLe-----GlnArgIleArgGlnGlnArgIleLeu 531
Db 30849 GTCCCTGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30908
QY 532 GlnArgLeuGlnLeuArgMetTyrLysLysGlyIleGlnGlnSerGlnProGlnVal 551
Db 30909 CACCGT-----CCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30937
QY 552 ThrSerPheProGlnProAspAspValGlnSerLeuMetIleThrProPheLeuPro 571
Db 30937 ----- 30937
QY 572 ValValAlaPheGlyArgProLeuProLysLeuThrProGlnAsnPheGlnLeuProTyr 591
Db 30938 -----CGACCGCGCGCGCTGAGAGACTGCGCAAGCGCGCTGAG--CCAGCA 30979
QY 592 LeuAsp-----GlnArgSerArgCysArgLeu 600
Db 30980 CGAGACCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 31012

```

RESULT 4
 US-09-320-878-19
 ; Sequence 19, Application US/09320878A
 ; Patent No. 6117659
 ; GENERAL INFORMATION:
 ; APPLICANT: ASHLEY, Gary
 ; APPLICANT: BETLACH, Melanie C.
 ; APPLICANT: BETLACH, Mary C.
 ; APPLICANT: MCDANIEL, Robert
 ; APPLICANT: TANG, Li
 ; TITLE OF INVENTION: RECOMBINANT MARBONOLIDE POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 300622002120
 ; CURRENT APPLICATION NUMBER: US/09/320, 878A
 ; EARLIER FILING DATE: 1999-05-27
 ; EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
 ; EARLIER FILING DATE: 1998-08-28
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
 ; EARLIER FILING DATE: 1998-05-06

Db	5353	GAA	CGGAGAGAGAGAGCGCCGACGAGCGGGAGAAAAAATATCCCGGAGAGAGAGAGTTCG	5412
Oy	251	Lys	SerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgMet----	268
Db	5413	CAG	CAGAGGAGGAAAGACAGTCTCTG--AGAGAGAGAACGGGAGAGAGAGAGCGCCAGAGAG	5469
Oy	269	-----	-----GlnLeuValLys	272
Db	5470	TGG	AGAGCGCAGTATCCCGAAAAAAGACAGCTGCACGAGAGAGAGAGAGAGAGCTGTGGA	5529
Oy	273	Lys	AspSerGlnGluArgHisLysLeuPheGlnGluArgGlu	292
Db	5530	GAG	AGAGACCGGAGAAAAAGAGA-----CTCCAGAGAGCGGAGAGGCAATATCCGGAG	5580
Oy	293	Glu	ThrGluLeuSerGlnLysIleLysLeuLysGlnProGluLeuSerGluThrSer	312
Db	5581	GAA	GAGAGCGTGCACAGC-----GAGAGAGAGCGCTGTGGAGAGAGAGCGG	5628
Oy	313	Gln	ThrLeuProProLysProPheSerCysGluArgSerGluLysGlnHisLysArgLys	332
Db	5629	GAG	AGC-----AGAAAGCGC	5643
Oy	333	Ser	ProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSer	352
Db	5544	CAG	AGAGCTGCAGAGGCAATATCGAAGAGAA-----GAGGAGCTGCACAGAGAGAGAG	5697
Oy	353	Lys	ValLysThrLysThrPro-----LysHisSerProIleLysGlnGluProCysGly	370
Db	5698	CAG	CTGCTGAGAGAGAGAACCGGAGAGAGAGAGCGCCAGGAGCGGAGAGCAATATCTCG	5757
Oy	371	Ser	LeuSerGluThrValLysLysArgGluLeuArgSerGlnGluThrProGluLysPro	390
Db	5758	GAG	AGAGAGAGCTGCACAGAGAGAGAGAGAGAGCAGCTGTGAGAGAGAGAAAGCGGAGAGAGA	5817
Oy	391	Arg	SerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHis	410
Db	5818	AGG	AGCGC-----CAGGAGCTGCAGAGGCAATAT	5844
Oy	411	Pro	LysGlnLysAlaPheSerSerGluIleGluAspLeuProLysLeuSerThrGlu	430
Db	5845	CGG	AGAGAGAGAGAGCTTACACCGCCAGAAAGAAACAGCATATCCGGAGTACAGATCAG	5904
Oy	431	Met	ThrLeuCysArgThrHisGlnProProProSerProLeuArgGluSerSer	450
Db	5905	CGC	AGTATCGAATATGGCAGCTGGAGAACAAAGAAATGCAATGTCGTGATTAACAG	5964
Oy	451	-----	-----ProLysGlnGluThrValAlaIleArgCysLeuMetProSerValAla	467
Db	5965	GTT	ATTCATCCAAAGGACAGAGAGATGAACAGTCCGGCAGTTGGAGAGATTCACAGTCCG	6024
Oy	468	Gly	LeuThrSer-----ValLeuAlaValProSerThrArgAspHis	481
Db	6025	GAC	AGACAAATCCAGCAAGATCTCAGACAGCTCTGGGTGAACAGAGAGAGATGCT	6084
Oy	482	Ser	ValGluProLeuArgAspProAsnProSer-----AspLeuLeu	495
Db	6085	GAG	CAAGAGAGAGCGCTGGCAGACAGGCCCAACAGGCAATTTCCAGAGAGAAACAGCTG	6144
Oy	496	Glu	AsnLeuAspAspSerValPheSerLysArgHisAlaLysLeu-----	510
Db	6145	GAG	CGAGAGAGCAAAAGAGAGCCAAAGCGCGCAGAGAGATCCCAAGAGAGAAAGAG	6204
Oy	511	-----	-----GluLeuAspGlnLysArgArgLysArgThrPaspIleGlnArgIleArg	526
Db	6205	TTC	CTGAGAGAGAAAGAGAGAGAGAGAGCGCTCAAGAGAGACAGAGAAATTCCTCG	6264
Oy	527	-----	-----GluGlnArgIleLeuGlnArgLeuGln-----LeuArgMetLysLys	541
Db	6265	GAG	AGAGAGAGAGCTGCTCTCAGAGAAAGGAGAGAACCGCTGCTCTCCCAAGAGCGTAC	6324
Oy	542	Lys	GlyIleGlnGluSerGluProGluValThrSerPhePheProGluProAspAspVal	561

Db	6325	AGAAATTCGGCAGAGGAAA-----	6345
Qy	562	GlusterleuNetIlethrProPhleuProIvalaIalaphleIyargProleuProlys	581
Db	6345	-----	6345
Qy	582	leuthrproGlnAsnPhleIuleuProTrIpleuAspGlnArgSerArgCysArgleuGlu	601
Db	6346	CTGCTCCATCAGGACAAGGAGAAAATTCTTCGAGAGAGAACAGCGCTGCGCAGGAAA	6405
Qy	602	IleGlnIyslys 605	
Db	6406	CGGAGAGGAAAA 6417	
RESULT 3			
US-09-105-537-5			
: Sequence 5, Application us/09105537A			
: Patent No. 6265202			
: GENERAL INFORMATION:			
: APPLICANT: Sherman, D.H.			
: APPLICANT: Liu, H.			
: APPLICANT: Xue, Y.			
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin			
: FILE REFERENCE: 600.438US1			
: CURRENT APPLICATION NUMBER: US/09/105.537A			
: CURRENT FILING DATE: 1998-06-26			
: NUMBER OF SEQ ID NOS: 43			
: SOFTWARE: FastSeq for Windows Version 3.0			
: SEQ ID NO 5			
: LENGTH: 36778			
: TYPE: DNA			
: ORGANISM: Streptomyces venezuelae			
US-09-105-537-5			
Alignment Scores:			
Pred. No.: 0.115 Length: 36778			
Score: 217.00 Matches: 177			
Percent Similarity: 31.13% Conservative: 63			
Best Local Similarity: 22.96% Mismatches: 256			
Query Match: 6.70% Indels: 282			
DB: 4 Gaps: 36			
US-10-054-935-2 (1-614) x US-09-105-537-5 (1-36778)			
Qy	1	MethrNetargSerAlaValaIphelysAlaIalAlaIalProAlaIagIylAsnProGlu 20	
Db	29019	GTCAACCTCGGCGCAAGTCCATTCGCGCCACCTCGCGGCAAGCGCGATGATCTCC 29078	
Qy	21	GlnArgleuAspTryGlnArgAlaIalAlaIeGlyGlyProGluAspGluProGly--- 39	
Db	29079	CTCGCCCTCGAC-----GAGGGGGCGCTCTTAAGCAGACTAGCGCACTTCGACGACTC 29132	
Qy	40	---AlaIlaGlnAlaIlaHisPhleuProArgHisArgLysLeuLysGluProGlyPro--- 57	
Db	29133	TTCGTGCGCGCGCGCAACG-CCCAACCGCGCACCGTGTCTCCGCGGACCGACCCAGAT 29191	
Qy	58	-----ProleuAlaSerSerGlnIglyly 65	
Db	29192	CGAAGACTCGCCCGCACTCGCAGGCGCGACGGCGTCCGCGCGATCATCCGGTCA 29251	
Qy	66	SerProAlaProSerProAlagly----- 73	
Db	29252	CTACGCTCCACACGCGCGCAGGTGCAGATCATCGAAGAGACTGGCGAGTCTCTCC 29311	
Qy	74	-----CysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaIylAlaIalProGly 90	
Db	29312	CGAGCTGCGCCCGGAGGCTCCGACAGCGGCGTCTTCT-CCACCTCGAAGGACACTGA 29370	
Qy	91	GlnGlnGlnGlu-----SerTrygly 97	
Db	29371	TCACCGAGCGGATGCTGACAGGCACTACTGTACCGCAACTGCGCATTCGGTGGCT 29430	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 20:15:46 ; Search time 80 Seconds

(without alignments)
2353.744 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238

Sequence: 1 MTMRSAVEKAAAPAGNPE.....RSRCRLTIQKQTHRTCRK 614

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_SPOOL/US10054935/runat_14072003_115025_24679/app.query.fasta_1.775
-DB=Issued_Patents_NA -QEXT=fastap -SUFFIX=rl -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORR=ext -HEAPSIZE=500 -MINDLEN=0 -MAXLEN=200000000
-USER=US10054935 @CGN 1.1.40 @runat_14072003_115025_24679 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	7.4	9551	1	US-08-056-200-93
2	238	7.4	9551	1	US-08-800-644-93
3	217	6.7	36778	4	US-09-105-537-5
4	217	6.7	38506	3	US-09-320-878-19
5	215.5	6.7	3147	2	US-08-781-802-7
6	215.5	6.7	3147	4	US-08-694-078-7
7	215.5	6.7	3147	4	US-09-058-260-7
8	214.5	6.6	4411529	4	US-09-103-840A-1
9	213.5	6.6	11958	4	US-09-134-246-8
10	212.5	6.6	4911	4	US-09-718-692-1
11	212.5	6.6	4911	4	US-09-718-852-1
12	212.5	6.6	4911	4	US-09-718-815-1

Result No.	Score	Query Match	Length	ID	Description
13	212	6.5	4403765	4	US-09-103-840A-2
14	211.5	6.5	2214	3	US-08-864-038A-1
15	211.5	6.5	3331	3	US-08-864-038A-2
16	211.5	6.5	3331	3	US-08-864-038A-4
17	210.5	6.5	33529	4	US-09-144-083-3
18	207	6.4	4403765	4	US-09-103-840A-2
19	206	6.4	3211	2	US-08-574-959A-8
20	206	6.4	3211	4	US-09-357-014-8
21	206	6.4	3901	2	US-08-574-959A-6
22	206	6.4	3901	4	US-09-357-014-6
23	206	6.4	13378	3	US-08-785-420-1
24	203	6.3	3283	4	US-09-651-656-16
25	203	6.3	3283	4	US-09-650-855-16
26	203	6.3	4411529	4	US-09-103-840A-1
27	201.5	6.2	2313	4	US-09-370-838-157
28	201.5	6.2	3239	4	US-08-927-219-5
29	201	6.2	43280	2	US-08-804-227C-1
30	200	6.2	4689	4	US-09-105-537-34
31	199.5	6.2	68750	3	US-09-335-409-1
32	199.5	6.2	68750	4	US-09-568-102-1
33	199.5	6.2	68750	4	US-09-567-969-1
34	199.5	6.2	68750	4	US-09-568-480-1
35	199.5	6.2	68750	4	US-09-568-486-1
36	199.5	6.2	68750	4	US-09-568-472-1
37	199.5	6.2	68750	4	US-09-567-899-1
38	199	6.1	1771	2	US-08-511-872-1
39	199	6.1	1771	2	US-08-533-669A-7
40	199	6.1	1771	4	US-09-183-861-7
41	199	6.1	1771	4	US-09-022-765-7
42	196.5	6.1	774	3	US-08-956-307B-12
43	196.5	6.1	778	3	US-08-956-307B-11
44	195.5	6.0	3132	2	US-08-224-482-3
45	195.5	6.0	3132	3	US-09-205-921-1

ALIGNMENTS

RESULT 1
US-08-056-200-93
Sequence 93, Application US/08056200
Patent No. 5616500

GENERAL INFORMATION:
APPLICANT: Steiner, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
SERIAL: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502

RL J. Biol. Chem. 274:32997-33001(1999).
 RN [6]
 RN INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
 RX PubMed:10433269;
 RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
 Doan A., Aakalu V.K., Lananan A.A., Sheng M., Worley P.F.;
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
 postsynaptic density proteins.";
 RL Neuron 23:583-592(1999).
 RN [7]
 RN INTERACTION WITH SPTAN1.
 RX PubMed:11509555;
 RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
 Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
 RT "Synaptic scaffolding proteins in rat brain. Ankryin repeats of the
 multidomain Shank protein family interact with the cytoskeletal
 protein alpha-fodrin.";
 RL J. Biol. Chem. 276:40104-40112(2001).
 RN [8]
 RN FUNCTION.
 RX PubMed:11498055;
 RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
 RT "Regulation of dendritic spine morphology and synaptic function by
 Shank and Homer.";
 RL Neuron 31:115-130(2001).
 RN [9]
 RN REVIEW.
 RX PubMed:10806096;
 RA Sheng M., Kim E.;
 RT "The Shank family of scaffold proteins.";
 RL J. Cell Sci. 113:1851-1856(2000).
 CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
 density (PSD) of excitatory synapses that interconnects receptors
 of the postsynaptic membrane including NMDA-type and metabotropic
 glutamate receptors, and the actin-based cytoskeleton. May play a
 role in the structural and functional organization of the
 dendritic spine and synaptic junction. Overexpression promotes
 maturation of dendritic spines and the enlargement of spine heads
 via its ability to recruit Homer to postsynaptic sites, and
 enhances presynaptic function.
 CC -1- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
 SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
 DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
 the PDZ domain (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
 neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
 CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms; 1, (shown here); 2, 3,
 4/5 and 5; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,
 CA1 region hippocampus and molecular layer of cerebellum).
 CC -1- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
 to high levels at 3-4 weeks before dropping slightly in adulthood.
 CC Expressed in the cortex and the molecular layer of the cerebellum
 at postnatal day 7. Isoform 2 expression does not change during
 development of both cortex and cerebellum. Isoform 4 expression
 decreases significantly during development of cortex but not
 cerebellum.
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 ANK REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF102855; AAD04569.2; -;
 DR EMBL: AF131951; AAD29417.1; ALT_INIT.

DR EMBL: AF159046; AAD42975.1; -;
 DR EMBL: AF141904; AAF02498.1; ALT_INIT.
 DR HSSP: P00519; IABL.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001660; SAM.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00023; ank; 7.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00536; SAM; 1.
 DR Pfam: PF00018; SH3; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00248; ANK; 3.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00454; SAM; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50086; ANK_REPEAT; 3.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR PROSITE: PS50105; SAM_DOMAIN; 1.
 DR ANK repeat: SH3 domain; Repeat; Alternative splicing.
 KW ANK repeat; 195
 FT REPEAT 195 210 ANK 1.
 FT REPEAT 212 245 ANK 2.
 FT REPEAT 246 278 ANK 3.
 FT REPEAT 279 312 ANK 4.
 FT REPEAT 313 345 ANK 5.
 FT REPEAT 346 378 ANK 6.
 FT REPEAT 379 395 ANK 7.
 FT REPEAT 395 413 SH3.
 FT DOMAIN 413 554 PDZ.
 FT DOMAIN 554 663 SAM.
 FT DOMAIN 663 757 SAM.
 FT DOMAIN 757 929 POLY-PRO.
 FT DOMAIN 929 1015 POLY-HIS.
 FT DOMAIN 1015 1027 POLY-HIS.
 FT DOMAIN 1027 1199 POLY-GLY.
 FT DOMAIN 1199 1850 POLY-PRO.
 FT VARSPPLIC 1850 613 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 613 654 SOEQROESRDKARLFPHYVYGSDFSDAPSLINDG
 > MALSAVGSGGSGALRDPPLALSSMPALGPRRSVWY
 1Y (IN ISOFORM 2).
 FT VARSPPLIC 646 654 MISSING (IN ISOFORM 3).
 FT VARSPPLIC 797 804 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 1930 1943 LSEDSOTSLSKPS -> QYRVVSSPDGDF (IN
 ISOFORM 5).
 FT VARSPPLIC 1944 2167 MISSING (IN ISOFORM 5).
 FT CONFLICT 1141 1141 S -> T (IN REF. 1).
 FT CONFLICT 1174 1174 S -> N (IN REF. 2).
 FT CONFLICT 1246 1246 R -> K (IN REF. 1).
 FT CONFLICT 1323 1323 A -> T (IN REF. 1).
 FT CONFLICT 1331 1331 S -> D (IN REF. 1).
 FT CONFLICT 1726 1726 S -> N (IN REF. 2).
 SQ SEQUENCE 2167 AA; 226333 MW; 3f478b5a7b18ba86 CRC64;
 Query Match 5.5%; Score 178.5; DB 1; Length 2167;
 Best Local Similarity 22.6%; Pred. No. 0.61;
 Matches 136; Conservative 48; Mismatches 211; Indels 207; Gaps 29;
 QY 16 GGNPQRDLRYERAAALGREDPEGAAEHFL-----PRHKKLEKRPPLASGSGSP----- 67
 DB 988 GGRGRKSLVH--SALLPRANNHRRPNNHNNHARRPNNHNNHARRP--EMETGSGSPDDRP 1044
 QY 68 -----APSPAGCGGGRGILLRPGA-APGQQRSEMSGVRLRCRP----- 106
 DB 1045 PRLALGPOPSLRGMWGGSPSPSGAPSPSHSSSGSGSPYQALRYFLRPPRAASAM 1104
 QY 107 -----PATKAGIGCEP-----AAAGACGSPR-----PKYQAVLPLOT 139
 DB 1105 YVPANSGRKRKPLVKQTKVEEERQKGSIPASSTPALRSEPPRPAEKKNSIRPT 1164
 QY 140 GSLVA-----AAKETPVNAGDKGCAASPA-----ATASDPAGPPPLP--- 178
 DB 1165 IIRKAPSTSSSGRSSSGSSTEAEPPTQPDGAGGCGGSGSPAPANTSPVPPSPSPVPTPASP 1224


```

OY 179 -GPEPLAPTAAGT-----LAASEGRL--KSMKRSPL-----GGGGSSGSSQAAC 221
DB 1225 SGPATLPTSGFALVGAARRREGGKMONERRSTLLSLDAGDEGDSGLGFGPPG- 1283
OY 222 LKQILLQLDLIEQOOOLQAKKEIELEKSERDTLLARIERMRRLQVKKDEKERHK 281
DB 1284 -----PRLR-HSKSIDGMEFSAEPLY-----RLSSGSSSGGYGVAAGSR 1322
OY 282 LFGYETEEREERELESEKIKLECOPELSETSQILPPKPFSCGSKCHKRRKSPG---S 337
DB 1323 AYGSGSSS-----SAFTSELPPLPVLPILGKGLDPAISGLALAA 1363
OY 338 TER-----KTPVKKIAPEFSKVTKPKKSP-----IEEPCGSLSEVCK 378
DB 1364 RERLAKSSGCGGTPPPPPPPSPRIADPPPTLHHHSHPSHPSHARHEPVLRIMGDPAR 1423
OY 379 RE-----LRSGE-----TPEKPRSY--DTPPR-----LSTPOKGPSTHPK----- 412
DB 1424 RELGYRAGLGSQEKALVAPSPAPARRSLHLRLPPAPGVPLQLGLGEPPTPHGVSKAW 1483
OY 413 EKAFSEIEDLPYLSTMYLCRMHQ-----PPSP---LPLRESSPKK 453
DB 1484 RTAAPEPERLP-LHVRFLNCOARPPAGTSGSSTEDGPGVPPSPRYLPTSPTRPG 1542
OY 454 EE 455
DB 1543 NE 1544

RESULT 12
RREL HUMAN STANDARD; PRT; 755 AA.
AC 092766;
DT 13-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RAS-responsive element binding protein 1 (RREB-1),
GN RREB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid carcinoma;
RX MEDLINE=96413283; PubMed=8816445;
RA Thilagalingam A., de Bustros A., Borges M., Jasli R., Compton D.,
RA Diamond L., Mabry M., Ball D.W., Baylin S.B., Nelkin B.D.;
RT "RREB-1, a novel zinc finger protein, is involved in
RT differentiation response to Ras in human medullary thyroid
RT carcinomas."
RL Mol. Cell. Biol. 16:5335-5345(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE
CC DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE
CC PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE
CC OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL
CC DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY AND PANCREAS. NOT FOUND IN THE BRAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U26914; AAB19094.1; -
CC HSSP; P08045; 12NF.
CC TRANSFAC; T01975; -

```

```

DR Genew; HGNC:10449; RREB1.
DR MIM; 602209; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 4.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS01057; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 315 337 C2H2-TYPE.
FT ZN_FING 461 483 C2H2-TYPE.
FT ZN_FING 580 602 C2H2-TYPE.
FT ZN_FING 608 630 C2H2-TYPE.
SQ SEQUENCE 755 AA; 79865 MW; 28D863EF62FE8415 CRC64;

Query Match 5.4%; Score 174; DB 1; Length 755;
Best Local Similarity 19.9%; Pred. No. 0.36;
Matches 142; Conservative 80; Mismatches 215; Indels 276; Gaps 32;

OY 34 PEDERGAEAHFLPRHKKLKEPPPLASSOGGSPAPSPA-----GGGKGRGLLPAGAA 88
DB 28 PEEEGSGSE-----QSP-----CPAPGPSLPVTLGPGILSESPAPAPAA 68
OY 89 -PGQOEESWGSGVPLPCP-----PPATKQAGIGEPBAAGAGC--SPRPYQAVLP 136
DB 69 TPPEAPQLOGPVOLANVPIYSALVSSPL-----VSSALLSGTALLRLPRKRPPLILP 123
OY 137 -----IOTGSLVAAA-----KEPP-----WAGDK--GGAASPAATA 166
DB 124 KPVTIELPLASTIAQIITSSVASPPLTKTKVADPGPASTGNTTASDSIGSVKAAAT 183
OY 167 SDPA-----GPPPLPLGPPPLATATA--GTLAASEGRKMSKSPSLGGGG----- 212
DB 184 ATPAATTSKRESSEPPAPASSPEASPTEGCAPATSKRKGRKMRSPRANGGVLDLS 243
OY 213 -----SCA----- 215
DB 244 SGERASTEKMLATDTTKFSPFLDTADNTODENAVGAPADHNGPSDEGSSPEDKLLRA 303
OY 216 --SSQAACLQIILLQLDLIEQOOOLQAKKEIELEKSERDTLLARIERMRRLQVKK 273
DB 304 KRNSYTNCLQIKTCPCPCPVFPVPMASLQHRXLTFTTDSQDAETFAAXGEVLD-----LTSR 359
OY 274 DNE-----KERKKLQGYETERETEISE----- 298
DB 360 DREQPSGATELROVAGDAVEQATATETASPVRHEEGRGESHPEEHEGTSTGADG 419
OY 299 -KILECOPE-----LSETSOTLPPKPFSCGSGRSGHRRKSPFGSTERKTPVKKLAPFSK 353
DB 420 GRGKVE-QPEPGRLRQAHGL-----QAGGRRGRGPRGGAASGE--QKLAQCTGC 467
OY 354 VKTKPRHSPKIEPCGSLSETVCKRELRSQETPEKPRSSVDPPLRSTPOKGPSTHPKE 413
DB 468 KSEKF-----LOTLSR--HRAKHROEPKDEKGDAT-----TAEEGSPAPEQ 509
OY 414 KAPSEIEIDLPLYSTMYLCRMHQPPSPPLRESSPKKEEYARCLM-----PSSVAG 468
DB 510 E-----EKPPETAEVESAAGAEAPAEKLAETEGPSD--G 545
OY 469 ETSVLAVSWRDSHVEPLRDPNPSDLLENDD-----SVFSK-----RHAKLELDE 514
DB 546 ESAAEKSSSEKSDDDKPKTDSKPSKASKADKKKVCYCNKRKWSIQLDTRHMRSTIGE 605
OY 515 KRRKRMIDIR-----IREQRILOQLDRMY-----KKKGIOSEBEVT 552
DB 606 RPYKCGCERCTFTLKSLVHORHIOKARAKHHGKSDKEEGEEDSENEST 658

RESULT 13
CSP_PLAKU STANDARD; PRT; 351 AA.
ID CSP_PLAKU
AC P04922;

```

DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Circumsporozoite protein precursor (CS).
 OS Plasmodium knowlesi (strain nurli).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5852;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85272582; PubMed=4023712;
 RA Sharma S., Spec P., Mitchell G.H., Godson G.N.;
 RT "Diversity of circumsporozoite antigen genes from two strains of the
 RL malaria parasite Plasmodium knowlesi.";
 CC Science 229:779-782(1985).
 CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
 CC VERTEBRATE HOST).
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M11031; AAA29540.1; -;
 DR PIR; A26253; OZ20KU.
 DR InterPro; IPR003067; Circmsprzoite.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF00090; TSP_1; 1.
 DR PRINTS; PR01303; CIRCMSPRZOITE.
 DR SMART; SM00209; TSP1; 1.
 KW Malaria; Sporozoite; Repeat; Signal.
 FT SIGNAL 1 19 PROBABLE.
 FT CHAIN 20 351 CIRCUMSPOROZOITE PROTEIN.
 FT DOMAIN 101 235 15 X 9 AA TANDEN REPEATS OF A-A-G-A-G-G-
 FT E-Q-P.
 SO SEQUENCE 351 AA; 34782 MW; A85E87A152E6485B CRC64;
 Query Match 5.3%; Score 173; DB 1; Length 351;
 Best Local Similarity 33.5%; Pred. No. 0.19;
 Matches 75; Conservative 12; Mismatches 95; Indels 42; Gaps 11;
 QY 16 GGNPEORDYERAAALGPEDEPGAAEAHFLPRHRKLKEPGPLASOGGSPAPSPACG 75
 DB 63 GEKPEGADKEK-----KEKEKEKEEPKPKNEKKLQPEPPAAGAGEGPA---AAG 114
 QY 76 GK---GKGLLPAGAPGQDESN-GGSVPLPCPPPTKQAGTIGEPAPAGAGCSPRPK 130
 DB 115 GEOPAAAGAGEOPAAAGAGEOPAAAGAGEOPAAAGAGEOPAAAGAG-GEOPA 173
 QY 131 YQAVLPITGSLVA-AKEPTPMAGDKG-----GAASPAATA--SDPA-----GPPPLP 176
 DB 174 AGA-----GGEPAPAGAGEOPAAAGAGEOPAAAGAGEOPAAAGAGEOPAP 228
 QY 177 LPPPPPLAPTAAGTILASEGKMKSRKSPILGGGSGASSQA 220
 DB 229 APRREGAPAGAVAGDGA-----RCGNAGAGGCGGCONNGA 263
 RESULT 14
 SFPQ_HUMAN STANDARD; PRT; 707 AA.
 AC P23246; P30808;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
 DE binding protein-associated splicing factor) (PTB-associated splicing
 DE factor) (PSF) (DNA-binding p52/P100 complex, 100 kDa subunit).
 GN SFPQ OR PSF.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal Brain;
 RX MEDLINE=93194059; PubMed=8449401;
 RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
 RT factor.";
 RL Genes Dev. 7:393-406(1993).
 RN [2]
 RP SEQUENCE OF 312-707 FROM N.A.
 RC TISSUE=Fetal skeletal muscle;
 RX MEDLINE=90091812; PubMed=2480877;
 RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
 RT "Cloning and characterization of a myoblast cell surface antigen
 RT defined by 24.1D5 monoclonal antibody.";
 RL Development 105:723-731(1989).
 RN [3]
 RP SEQUENCE OF 48-68 AND 213-246.
 RX MEDLINE=93176127; PubMed=8439294;
 RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 RT "Purification and characterization of a DNA-binding heterodimer of 52
 RT and 100 kDa from HeLa cells.";
 RL Biochem. J. 290:267-272(1993).
 CC -1- FUNCTION: ESSENTIAL. PRE-mRNA SPLICING FACTOR REQUIRED EARLY IN
 CC SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
 CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
 CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
 CC -1- SUBUNIT: HETEROETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
 CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
 CC ECTOKINASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X70944; CAA50283.1; -;
 DR EMBL; X16850; CAA34747.1; -;
 DR PIR; A43557; A43557.
 DR PIR; S29770; S29770.
 DR HSSP; P11940; 1CVY.
 DR SWISS-2DPAGE; P23246; HUMAN.
 DR Genew; HGNC:10774; SFPQ.
 DR MIM; 605199; -;
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS50102; RRM; 2.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
 KW Alternative splicing.
 FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
 FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
 FT REPEAT 9 11 3 X 3 AA REPEATS OF R-G-G.
 FT REPEAT 19 21 2.
 FT REPEAT 25 27 3.

```

FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPLIC 663 707 RTERFGGAGPVGGGPRGMPETPACRGREYRG
FT PNRKPR -> VRAIDVG (IN SHORT ISOFORM).
FT CONFLICT 243 243 G -> R (IN REF. 3).
SQ SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;

Query Match 5.3% Score 172.5; DB 1; Length 707;
Best Local Similarity 22.3%; Pred. No. 0.39; 200; Indels 219; Gaps 32;
Matches 137; Conservative 58; Mismatches 200;

QY 49 HRRLKEPGRPLAS-----QGSPAPSPAGCGKRGILLPA-----GAAPGQ 91
D 94 HQQQPPPPPPQDSSKPVVAGGPAP-----GVGSAPPASSAPAPPTPGAPG- 144
QY 92 QESWGSVPLPCPPRATKAGIGGERPAAGGSPRKYQAVLPIDGSLVAAAKPT- 150
D 145 -----SGPPTPPPPVTSAPRAPP-----PPSSSGV-----PTT 177
QY 151 -PMAGDGAASAPATSDPAPRPLPGRP-----PLPTATAGTLAASGKWKMRK 204
D 178 PPOAG--GPPPPPAVNGPGRPGKGTGPGKMGKGGKGGGGLSTPG--GHPK 232
QY 205 SPLGGGSGASSQAACILQLDLLEQOQO-----LQAKEIEELKSBDT 256
D 233 PPRHGGGERPGRQ-----NHPRYQDHHQGRPGGGRSEK----- 271
QY 257 LLAIEEMERMOVKQDNEK---ERKILFOG---YETEREETELSEK----- 299
D 272 -ISDSEGFKANLSLRPRGERTYQRCRLFVGLPAITDEEFRLKAYGEPGEVINK 330
QY 300 -----IKLECPRELSTQTLPPK-----FSCGRSGKHKRKS----- 334
D 331 GKGGFKLESRALAETAKAELDTPRGRQLRYRFTHAALSVRNLSPVSNELLEEA 390
QY 335 ---FGSTER-----KTPYKLAPEFSK---VKTKPKHSPIKEE 367
D 391 FSGEGPIERAVVIYDGRSGTGKIVFASKPAARKAFERCSBEVFLTTTPR--PVIYE 448
QY 368 PCGSLSTVCKRELRSQETP--EKPNSSVTPPRLSTPQKGPSTHPEKAPSSIEDLPY 425
D 449 PLEQLDDEDLPEKLAQKPMYOKER---ETPPFA---QHG-----TFEYE----- 489
QY 426 LSTTEMLCRHWDPPPSPLPLRESSPKETVARCLMPSSVAGETSVLAVPSMDHSEV 485
D 490 -----YSQWKKS-----LDEMEKQOREOVENMDADOKLESEN--EDAEHQAHL 534
QY 486 LRDPNPSDL-----LENLDVSFSKR--NAKLEDEKRRKMDIQIRIORTI--LQ 532
D 535 LR-----QDLMRQBELRMELINQEMQKKEKQLRQEEERRRREEMLRQEMEDMR 590
QY 533 RLQLRMVKKKGIOE 546
D 591 ROREESYSRMGYMD 604

RESULT 15
HRX_HUMAN STANDARD; PRT; 3969 AA.
AC 003164; Q14845; Q16364; Q13743; Q13744; Q9UM43;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (ALL-1) (Trithorax-like protein).

```

```

GN MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93046667; PubMed=1423624;
RA Trachuk D.C., Kohler S., Cleary M.L.;
RT "Involvement of a homolog of Drosophila trithorax by 11q23
RT chromosomal translocations in acute leukemias.";
RL Cell 71:691-700(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9629053; PubMed=8703835;
RA Nilsson I., Loechner K., Siegler G., Grell J., Beck J.D., Fey G.H.,
RA Marschalek R., Sieglar G., Grell J., Beck J.D., Fey G.H.,
RT "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT translocations to chromosomal region 11q23 and acute leukemias.";
RL Br. J. Haematol. 93:966-972(1996).
RN [3]
RP SEQUENCE OF 1-1909 FROM N.A.
RX MEDLINE=93390935; PubMed=8378076;
RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT "Two distinct portions of LTR19/ENL at 19p13 are involved in t(11;19)
RT leukemia.";
RL Oncogene 8:2617-2625(1993).
RN [4]
RP SEQUENCE OF 1317-2328 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93265134; PubMed=1303259;
RA Diabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT "A trithorax-like gene is interrupted by chromosome 11q23
RT translocations in acute leukemias.";
RL Nat. Genet. 2:113-118(1992).
RN [5]
RP SEQUENCE OF 1251-1538 FROM N.A.
RX MEDLINE=94215165; PubMed=8162575;
RA Gu Y., Alder H., Nakamura T., Schlichman S.A., Prasad R., Cnaanl O.,
RA Saito H., Croce C.M., Cnaanl E.;
RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT involved in acute leukemia.";
RL Cancer Res. 54:2326-2330(1994).
RN [6]
RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX MEDLINE=95322025; PubMed=7598802;
RA Manganolli D., Burnett R., McCabe N., Thitman M., Gall H., Yu H.,
RA Rowley J.D., Diaz M.O.;
RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT trithorax-like domain, and alternative splicing.";
RL DNA Cell Biol. 14:475-483(1995).
RN [7]
RP SEQUENCE OF 1212-1603 FROM N.A.
RX MEDLINE=95315013; PubMed=7794749;
RA Marschalek R., Grell J., Loechner K., Nilsson I., Siegler G.,
RA Zweckroner I., Beck J.D., Fey G.H.;
RT "Molecular analysis of the chromosomal breakpoint and fusion
RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT translocation t(4;11).";
RL Br. J. Haematol. 90:308-320(1995).
RN [8]
RP SEQUENCE OF 1421-1540 FROM N.A.
RX MEDLINE=94020842; PubMed=841518;
RA Forster A., Rabbits T.H.;
RT "A method for identifying genes within yeast artificial chromosomes:
RT application to isolation of MLL fusion cDNAs from acute leukemia
RT translocations.";
RL Oncogene 8:3157-3160(1993).
RN [9]
RP CHROMOSOMAL TRANSLOCATION WITH GAST.
RX MEDLINE=20183971; PubMed=10706619;
RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,

```



```
QY 331 -----RSPFGSTERKTPVK-----KLAPFSKVKT-----KTPKH 361
Db 250 EDSLKIKIKTPPSATFQOATKIKIKLACKLSPLKSKFKTKGLQIGRKGVQIYRRGRPPST 309
QY 362 SPIKEEPCGSL--SETVCKRELRSQETPEKPRSSVD-----TPPRLSTPOKGPSTH--- 410
Db 310 ERKIK-TPSGLLINSELEKPOKVRKDKEGTPELTKEKTVVRQSPRRKIKPVRIIPSSKRTD 368
QY 411 -----PKKAFSSSEIEDLPYLS 427
Db 369 ATIAKOLLQRAKKAQAKKIEKEAAQLQGRKVTQVKNIROFIMPVYSAISSRI-----IK 423
QY 428 TTEMVLCRMHQPPSPPLPLRESSPKKEETVARCLMSSVAGETSVLAVPSWRDHSVEPLR 487
Db 424 TPRRFIEDEDYDPPIKIARLESTPN-----SRESAPSCGSSSEKSSAASQHSQSSSDSSR 478
QY 488 DPNPSDLLENLUDSVFSKR-----HAKLELDEKRRKRMIDIORIREQRIOR 533
Db 479 SSSPS--VDRSTDSQASEIQLVLPERSDTPVEVHPPLPISQSPENESNDRSRKRYSVSER 536
QY 534 -IQLRMYKKKKIQESEP-EVTSFEPEDVESLMTPELPVAFGRPLPKLTPQNF---E 588
Db 537 SFGSRFTTKLSTLQSAPOQOTSSSPP-----PLT-----TPPPPLQPASSISDH 581
QY 589 LPWL 592
Db 582 TPWL 585
```

Search completed: July 14, 2003, 11:49:04
Job time : 29 secs

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED
1	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning	Mech. Enzymol.	303, 19-44 (1999)	99279253
2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	Genome Res.	10 (10), 1617-1630 (2000)	10349636
3	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., Nishikawa, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiyagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer	Genome Res.	10 (11), 1757-1771 (2000)	11042159
4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Atzawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, K., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochava, H., Knehl, P., Lewis, S., Matsuo, Y., Stakalo, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabile, F., Suzuki, R., Tomita, M., Wagner, T., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Beffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotamian, M., Hume, D. A., Kamlay, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombereis, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyono, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmung, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, I., Kohlsuki, S. and Hayashizaki, Y.	Functional annotation of a full-length mouse cDNA collection	Nature	409 (6821), 685-690 (2001)	21085660
5	(bases 1 to 3161)				11217851
REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED
1	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F., Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasai, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, K. I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	Direct Submission			
2	Submitted (10-JUL-2000)	Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome_research@riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,			

COMMENT

Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research at Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGACGAAGATCCAGAAGCTCTTTTTCCTTTTTTAA 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer abap of sequence [5' GAGAGGAGCTGTCCAGTAATAAATTATATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size compressed longer than 7 kb was selected before cloning. Vector: a modified pluciscrypt KS(+), after bulk excision from Lambda FLIC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source	location/Qualifiers
	1..3161
	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="FANTOM_DB:4121402D02"
	/db_xref="MGD:MGI:1907076"
	/db_xref="taxon:10090"
	/clone="4121402D02"
	/tissue_type="head"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="16 days embryo"
	316..2166
	/note="PRICHTEXTNSN domain containing protein data source:InterPro, source key:IPR002965, evidence:ISS putative"
	/codon_start=1
	/protein_id="BAB29369.1"
	/db_xref="GI:12852333"
	/db_xref="MGD:MGI:1921276"
	/translation="MTMRSAYFKRAASAPAGCNPNROLDYERAAALPGQEDSGAAEAH FLPHRLKKEPPLASSOGSSPPSPAGCGGRGLILPAGAAAPQOESMSGSVP LPCGPATKGIGIEPVNAAGCSPPPKTOAVLPITGSTVVAAREPTPMADCKGG AAPAAATASDPAGPPPLPLPGPPPLPATATGTAASEGRKSIRKSPGLGGGSGAS SOAKLKIOILLDLLEOQQLOAKKELEELSESDTLARIERMERMOVLKRD NEKRHKLLQGEPEEREAEALSEKIILERPELCETSQALESFGSRGSKHKRRK TPRENTKRNKYKLAFEPBSFVKTKTKRHSPKIREKPGSISETVCRELROEFPKRR RSSVDTPRLSTPOKGFSTHRKEAFSEEMDDPLYLTETTELDRNHQPSPSLPIARR SSPFKEELVARCLMPSSVAGETSVLAIVPSWDSHVPELRDNPSDILENDVSFSRK HAKLEDEKRRKMDIQRIEORILORLOLMRYKKKIOISEPEVTSFFPDVESEL LITFLPVLVAFGRLPLKPAONFELPWMLDEERSRCRLIOKKKHPTHCRK"

CDS

BASE COUNT 721 a 855 c 911 g 674 t

ORIGIN

Alignment Scores:

Pred. No.:	3,32e-179	Length:	3161
Score:	3110.00	Matches:	591
Percent Similarity:	97.56%	Conservative:	10
Best Local Similarity:	95.94%	Mismatch:	13
Query Match:	96.05%	Indels:	2
DB:	11	Gaps:	2

US-10-054-935-2 (1-614) x AK014463 (1-3161)

OY	1	MethrhetarqserAlavalphelysAlalalaalaproialaglyglAsnProglu	20
Dd	316	ATGACCATGAATCCGCAGTGTTCAAGCGGCCCGGCCTTCGCGGGCAACCCGAG	375
OY	21	GlmarqleuaapyrgluArAlalalaleuglygPrugiuaasgluPrOglyVala	40
Dd	376	CAGGACTGCGACTACGAGCGGGCTGGCGGCGGGCCGAGAGAGAGATCCGGGGCG	435
OY	41	AlaqlualahIsphaleuProArGHsarqylsleuysgluPrOglyProProLeuala	60
Dd	436	GCCGAAGCCATTTCTCCGCCCGGCATGTGAAGTCAAGAGAGCGGGGGCCCGCTGGCC	495
OY	61	SerSerlngllyglierProialaProSerProialaclys--GlyglyLysglyArg	79

Db	496	TCTTCCAGGCGGGAGACCCCTCCCTCTCCACCCGCTCGCGCGGCGGAAGGCGCGG	555
Qy	80	GIyleuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnSerTrpGlyGlySer	99
Db	556	GGCTGTATTACCTCCAGCGGGGGGGCCCGCGGACGAGAAAGAGACTGGGGCGTTCC	615
Qy	100	ValProLeuProCysProProProAlaThrLysGlnAlaGlyIleGlyGlyGluProAla	119
Db	616	GTGGCCTTATGGCTCTCCGCGCCCGAGCTAACTAAACGGCGCATTCGGGGGAGCAACATC	675
Qy	120	AlaAlaGlyAlaGlyCysSerProArgProLysTrpGlnAlaValLeuProIleGlnThr	139
Db	676	GCACCCGGCGCTGGCTGCAGCCCGCGGCCCAAGATCAGGGGGTGGCTCCCATTCAGACG	735
Qy	140	GlySerLeu---ValAlaAlaAlaLysGluProThrProTPAlaGlyAspLysGlyGly	158
Db	736	GGCTCTATCGGGTGGGGGGCGGCCAAGAGGCTTACGCTTGCGCTGGGGAGCAAGGGTGG	795
Qy	159	AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro	178
Db	796	GGGCTCCCGCCAGCTGCACCGCTCGGAGCCCGCGGGAGCCCGACACCTACCTCGGCC	855
Qy	179	GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyLysArg	198
Db	856	GGGCGACACCCCTCGCGCGCCACCGCCACCTGTGGACCCGCGGGCGCAGTGAAGGCGCA	915
Qy	199	TrpLysSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGln	218
Db	916	TGGAGAGATTAAGAGAAAGAGCCCTCTCGGGGGTGGCGGGCGCTGGGAGCCTCCAGTCA	975
Qy	219	AlaAlaCysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGln	238
Db	976	GCCCGCTAGCCTCAAAACAGATCCTTCTGCTGCAATTGGACCTCATTCGAACAGACAGAG	1035
Qy	239	GlnLeuGlnAlaLysGluLysGluIleGlnGlnLeuLysSerGluArgAspThrLeuLeu	258
Db	1036	CAGTTGGACGGCAAGAGAAAGAAATGAGAGAGCTGAAGTCCGAGAGAGAACTCCTT	1095
Qy	259	AlaArgIleGluArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluGluGlu	278
Db	1096	GCTCGGATGTAGACTATGAGAAAGCGGATGCACTGGTGGAAAGAGGATTAACGAGAAAGAA	1155
Qy	279	ArgHisLysLeuPheGlnGlyTyrGlnThrGlnGluArgGlnGluThrGluLeuSerGlu	298
Db	1156	AGGCACAAGCTGCTCCAGGGCTATGAACCTGAAGAGAGAGAGAAAGCAAGATTGTCTAG	1215
Qy	299	LysIleLysLeuGluCysGlnProGluLeuSerGlnThrSerGlnThrLeuProProLys	318
Db	1216	AAAAATTAATTTGGAGCGCGCACCGGAGCTTTGGCAGACCTCCAGCGCTCTCCCTCCAAAG	1275
Qy	319	ProPheSerCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThr	338
Db	1276	CCTTCTCATGTGGCGGGAGTGAAGAGGACACAAAGAGAAACCCCATTTGGAAATTCACA	1335
Qy	339	GluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThr	358
Db	1336	GAAGAAAGAAATCCGTAAAAACCTGGCTCTGTAATTTCAAAAGTCAAAACAAAAACT	1395
Qy	359	ProLysHisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCysLys	378
Db	1396	CCTAAGCACTCCCATTTAAAGAGAACCTGTGGTTCATATCAGAAACGTGTGTAA	1455
Qy	379	ArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProPro	398
Db	1456	CGTAAATTTGAGAGACCAAGAAACCCCAAGAAAGCCCGGCTTCAGTGAATACCCACACA	1515
Qy	399	ArgLeuSerThrProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSer	418
Db	1516	AGACTCTTGACTCCCAAAAGAGAACCGACACCCACCCAGAGAAAGCCTTCTCAAGT	1575
Qy	419	GluIleGluAspLeuProTyrLeuSerThrThrGlnMetTyrLeuCysArgTrpHisGln	438

Db		1576	GAGATGGAAGATTGGCCGTACCTTTCACCACACACAATAAATGTAATTGTGTGCGTTGGCACGAC	1635
Oy		439	ProPProSerProLeuProLeuArgLusSerProIysLysGLuThrValAla	458
Db		1636	CCTGCCCATCACCGETTACCACTATTCAGGAAATCTCTCCAAAGAAGAGAGAACACTGTAGCA	1695
Oy		459	ArgCysLeuMetProSerSerValAlaGlyLutThrSerValLeuAlaValProSetTrp	478
Db		1696	AGGCTCTGAGATCCATCACTAGTGTTGCAGAGAAAACCTCAGCTTGGCTGTCTTCCTTTCGG	1755
Oy		479	ArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuAsnLeu	498
Db		1756	AGGGACCACTCAGTAGACCTCTTAAGGACCCCAAAATCCTTGACATTTTGGAGAACCTG	1815
Oy		499	AspAspSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLysArgArgLys	518
Db		1816	GATGACACTGATATTTTCAAAGAGCGATGCCAAAACTGGAGCTTAGACAGAAAGAGAAAA	1875
Oy		519	ArgTrpaspIIleGlnArgIIleArgGluGlnArgIIleLeuGlnArgLeuGlnLeuArgmet	538
Db		1876	CGATGGGATATTTCAGAGAGATCAGGAGAACAAAGAATTTTACAGCACCTGCACCTCAGATG	1935
Oy		539	TyrIlysLysLysGlyIIleGlnGlnSerGluProGluValI ThrSerPhePheProGluPro	558
Db		1936	TATATAAAGAAAGAAATTCAGAAATCTAGCCTGAGGTGTACTCTATTTTCCCTGAGGCA	1995
Oy		559	AspAspValIGlueSerLeuMetCilethrProPheLeuProValAlaIaPheGlyArgPro	578
Db		1996	GATGATGTTGAAAGTTTCTGCTGATNACCCCCTTCTTGCTGTGTAGACATTTGGAGCGCCA	2055
Oy		579	LeuProLysLeuThrProGlnAsnPheGluLeuLeuPotrLeuAspGluArgSerArgCys	598
Db		2056	TTGGCCAAATATTAGCTCCACAGAACCTTGMACTACCTGTTGGATGAGCGAAGCCGTGTGC	2115
Oy		599	ArgLeuGluIIleGlnLysLysGlnThrProHisArgThrCysArgLys	614
Db		2116	AGGTTGGAGATCCAGAAAGACACACACTCTACCGGACCTGTAGGAG	2163
RESULT 2				
AKO15496	LOCUS			
DEFINITION		AKO15496	1304 bp mRNA linear HTC 19-JAN-2002	
		Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930463F05:hypothetical protein, full insert sequence.		
ACCESSION		AKO15496		
VERSION		AKO15496.1	GI:12853861	
KEYWORDS		HTC; CAP trapper;		
SOURCE		Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library		
		clone:4930463F05.		
ORGANISM		Mus musculus		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE				
AUTHORS		1		
TITLE		Carninci,P. and Hayashizaki,Y.		
JOURNAL MEDLINE		High-efficiency full-length cDNA cloning		
PUBMED		Meeth. Enzymol. 303, 19-44 (1999)		
99279253				
10349636				
REFERENCE		2		
AUTHORS		Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL MEDLINE		Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED		20499374		
11042159				
REFERENCE		3		
AUTHORS		Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitounai,T., Tashiro,H., Itoh,M., Suno,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Itawa,M., Ohara,E., Watabiki,M.,		

OY 406 GLYProSerThrHisProLysGluValAlaPheSerSerGluIleGluAspLeuProTyr 425
 |||||||
 DB 548 GGACCACACACCCACCCAGAGAAAGCCCTTCTCAAGTGAATGGAAGATTGGCGTAC 607
 OY 426 LeuSerThrThrGluMetTyrLeuGlyArgTyrPheIleGlnProProProSerProLeuPro 445
 |||||||
 DB 608 CTTTCACACAGAAATGTAATTGTGTGGTGGACCAAGCCGCCCATTCACCGTTACCA 667
 OY 446 LeuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSer 465
 |||||||
 DB 668 TTACGGGAATCCTCTCCAAAGAGAGAGACTGTA----- 703
 OY 466 ValAlaGluThrSerValLeuAlaValProSerTyrPArgAspHisSerValGluPro 485
 |||||||
 DB 704 -----GCATTCCTCTTTGGAGGACCACTCAAGTACGCT 739
 OY 486 LeuArgAspProAsnProSerAspLeuGluGluAsnLeuAspAspSerValPheSerTyr 505
 |||||||
 DB 740 CTAAAGGACCCAAATCCTTCAGACATTTTGGAAAGCTGGATGACAGTATTTCAAG 799
 OY 506 ArgHisAlaLysLeuGluLeuAspGluLysArgTyrGlySarGlyTyrPaspIleGlnArgIle 525
 |||||||
 DB 800 AGCGATGCAAAAGCTGAGCTAGACAGAGAGAGAGAAACGATGGATATTCAGAGATC 859
 OY 526 ArgGluGlnArgIleLeuGluArgLeuGlnLeuArgMetTyrLysLysLysGlyIleGln 545
 |||||||
 DB 860 AGGGAACAAGAAATTTTACAGCGACTGCAGCTCAGAAATGTATTAAGAAAGAAATTCAG 919
 OY 546 GluSerGluProGluValThrSerPhePheProGluProAspAspValGluSerLeuMet 565
 |||||||
 DB 920 GAATCTGAGCCTGAGAGTTACCTCATTTTCCCTGAGCAGATGATGTAAGTTGGCTG 979
 OY 566 IleThrProPheLeuProValValAlaPheGlyArgProLeuProLysLeuThrProGln 585
 |||||||
 DB 980 ATTACCCCTCTCTGCTGTTGATGATTCAGCGCGCATTCGCAAAATTCCTCCACAG 1039
 OY 586 AsnPheGluLeuProTyrPheAspGluArgSerArgCysArgLeuGluIleGlnLys 605
 |||||||
 DB 1040 AACTTGAACACTACCTGTTGATGATGAGCGAAGCCCTTCAGAGTTGAGATCCAGAGAG 1099
 OY 606 GlnThrProHisArgThrCysArgLys 614
 |||||||
 DB 1100 CACACACCTCACCGGACCTGTAGAGAG 1126
 RESULT 3
 B0014717/c 765 bp mRNA linear EST 26-MAR-2002
 LOCUS UI-H-ED1-ext-j-04-0-UI.s1 NCI_CGAP_ED1 Homo sapiens cDNA clone
 DEFINITION IMAGE:5833443 3', mRNA sequence.
 B0014717
 ACCESSION B0014717.1 GI:19739618
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 765)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Seq primer: M13 FORWARD
 Polya-yes.
 FEATURES Location/Qualifiers

source 1..765
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5833443"
 /clone_lib="NCI_CGAP_ED1"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Left Pubic Bone; Vector: pT73-Pac
 (Pharmacia) with a modified polylinker; Site:1: EcoR I;
 Site:2: Not I; NCI_CGAP_ED1 is a normalized cDNA library
 containing the following tissue(s): Chondrosarcoma cell
 line C55. The library was constructed according to Bonaldo
 /Lennon and Soares, Genome Research, 6:791-806, 1996.
 First strand cDNA synthesis was primed with an oligo-dT
 primer containing a Not I site. Double stranded cDNA was
 ligated to an EcoR I adaptor, digested with Not I, and
 cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is GCTCAAGCT.
 TAG_Lib=UI-H-ED1
 TAG_TISSUE="Chondrosarcoma"
 TAG_SEQ="GCTCAAGCT"
 BASE COUNT 158 a 176 c 180 g 250 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,85e-71 Length: 765
 Score: 1324.00 Matches: 251
 Percent Similarity: 99.60% Conservative: 0
 Best Local Similarity: 99.60% Mismatches: 1
 Query Match: 40.89% Indels: 0
 DB: 14 Gaps: 0
 US-10-054-935-2 (1-614) x B0014717 (1-765)
 OY 291 ArgGluGluThrGluLeuSerGluLysIleLysLeuGluGlnProGluLeuSerGlu 310
 |||||||
 DB 763 AGAGAGGAAGACGACTATCTGAGAAATTAACGAGAGTGCACCGAGCTTCCGAG 704
 OY 311 ThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyHisLys 330
 |||||||
 DB 703 ACATCCAGACCTCTCCCTCCCAAGCCCTTCTATGTGGCGGAGGAGGAAGACATATAA 644
 OY 331 ArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGlu 350
 |||||||
 DB 643 AGGAATATCCCATTTTGAAGTACAGAAAGAAAGAACTCTGTTAAAGACTGCTCTGAA 584
 OY 351 PheSerLysValLysThrLysThrProLysHisSerProIleLysGluGluProCysGly 370
 |||||||
 DB 563 TTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCATTTAAAGAGAACCTGTGAT 524
 OY 371 SerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysPro 390
 |||||||
 DB 523 TCCTTATCTGAAACACTGTTGTAAAGCTGATGAGAGACCAAGAAACCCAGAAAGGCC 464
 OY 391 ArgSerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHis 410
 |||||||
 DB 463 CGGCTTCAGTGGACACCCACCAAGACTCTCCATCCCAAAAGGAGCCAGCACCAT 404
 OY 411 ProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrArgu 430
 |||||||
 DB 403 CCCAAGGAGAAAGCCCTTCAAGTAGAGTAAGAAATTTGGCCGTACCTTCCACCAAGAA 344
 OY 431 MetTyrLeuCysArgTyrPheIleGlnProProProSerProLeuArgLysSerSer 450
 |||||||
 DB 343 ATGTATTGTGTGCTGGCACCAAGCTCCCATCAGTACATTCAGGAAATCTCT 284
 OY 451 ProLysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyLysThr 470
 |||||||
 DB 283 CCAAGAGAGAGAGAGAGTGTAGCAAGGTGTGATGCATCAAGTGTTCAGAGAGAAACT 224

QY 471 SerValLeuAlaValProSerTrpArgAspHisSerValGluProLeuArgAspProAsn 490
Db 223 TCACCTCTGGCTGCTTCCTTCCTTGGAGGACACTAGAGCCTTAAGGAGCAACCAAT 164
QY 491 ProSerAspLeuLeuGluAsnLeuAspSerValAlaPheSerLysArgHisAlaLysLeu 510
Db 163 CTTTCACACCTTTTGGAGAACCTGATGACAGTGTGTTTCGAAGCGCATGCAAAACTG 104
QY 511 GluLeuAspLysArgArgLysArgTrpAspIleGlnArgLysLeuArgLysLeu 530
Db 103 GAGCTGATGAGAGAGAGAGAAAGATGGATATTCAGAGCATGAGGAGAAAGAAATT 44
QY 531 LeuGlnArgLeuGlnLeuArgMetTrpLysLysLys 542
Db 43 TTACAGCAGCTGCAGCTCAGAAATGATATAAAAAA 8
RESULT 4
LOCUS BQ229745 870 bp mRNA linear EST 02-May-2002
DEFINITION AGENCOURT_7510022 NIH_MGC_92 Homo sapiens CDNA clone IMAGE:6042882
ACCESSION BQ229745
VERSION BQ229745.1 GI:20411145
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 870)
JOURNAL NIH-MGC http://mgc.ncl.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
plate: LLAM13282 row: P column: 19
High quality sequence stop: 596.
FEATURES
Source
Location/Qualifiers
1. 870.
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6042882"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 171 a 289 c 302 g 107 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7.09e-70 Length: 870
Score: 1302.00 Matches: 260
Percent Similarity: 92.07% Conservative: 7
Best Local Similarity: 89.66% Mismatches: 14
Query Match: 40.21% Indels: 9
DB: 14 Gaps: 1
US-10-054-935-2 (1-614) x BQ229745 (1-870)
QY 13 AlaProAlaGlyLysAnProGluGlnArgLeuAspTrpGluArgAlaAlaLeuGly 32
Db 2 GCCCCTGCCGCGGCGGCAATCTCTGAGCAGCGACTGACACGAGCGGCTGCGCGCTGGCC 61

QY 33 GlyProGluAspGluProGlyAlaAlaHisPheLeuProArgHisArgLysLeu 52
Db 62 GGGCCGAG 121
QY 53 LysLeuProGlyProProLeuAlaAlaSerSerGlnGlyLysProAlaProSerProAla 72
Db 122 AAG 181
QY 73 GlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGln 92
Db 182 GCGTGGCGGCGGAG 241
QY 93 GluGlnSerTrpGlyGlySerValProLeuProCysProProProAlaTrpLysGlnAla 112
Db 242 GAAG 301
QY 113 GlyIleGlyGlyGluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTrpGln 132
Db 302 GCGATTTGGGGGGGAG 361
QY 133 AlaValLeuProIleGlnTrpGlySerLeuValAlaAlaLysGluProTrpProTrp 152
Db 362 GCGGTGCTGCCATTCAG 421
QY 153 AlaGlyAspLysGlyLysAlaAlaAlaSerProAlaAlaTrpAlaSerAspProAlaGlyPro 172
Db 422 GCTGGGAGACAAG 481
QY 173 ProProLeuProLeuProGlyProProProProProLeuAlaProTrpAlaTrpAlaGlyTrpLeu 192
Db 482 CCACCACTACTCTGCCCCGGGGGGGGCCACCCCTGCGCCAGCCGAGAGAGAGAGAGAGAGAG 541
QY 193 AlaAlaSerGlnGlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlyGly 212
Db 542 GCGGCGGCGGAG 601
QY 213 SerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeu 232
Db 602 TCGGAGAGCTTCAGCTGAGCGCGCGCTGCTCAACAGATCCTTCTGCTGCAATTGGAGCTC 661
QY 233 IleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 252
Db 662 ATGACACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 721
QY 253 -GluArgAspTrpLeuLeu-AlaArgIleGlu-ArgMetGluArgArgMetGlnLeuVal 271
Db 722 NAAAG 781
QY 272 LysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTrp----- 286
Db 782 GTAAG 841
QY 287 ---GluTrpGluGluArgGluGlu 293
Db 842 TTAAG 865
RESULT 5
LOCUS B1665707 854 bp mRNA linear EST 12-SEP-2001
DEFINITION 603287711F1 NCL_CGAP_Mam6 Mus musculus CDNA clone IMAGE:5321855 5',
mRNA sequence.
ACCESSION B1665707
VERSION B1665707.1 GI:15579940
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 854)
JOURNAL NIH-MGC http://mgc.ncl.nih.gov/
Unpublished (1999) National Institutes of Health, Mammalian Gene Collection (MGC)

QY 130 LysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaValGluPro 149
 DB 361 AAGTATCAGGGGGTGTGCCCATTCAGACGGGGCTCTCGTGGGGGGCCAAAGAGCT 420
 QY 150 ThrProTrrPalaglyAspLysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspPro 169
 DB 421 ACGCCCTGGGCTGGGGACAAAGGTGGGGCGCTCCCGAGCTGCACCGCTCGACCGG 480
 QY 170 AlaGlyProProLeuProLeuProGlyProProLeuAlaProThrAlaThrAla 189
 DB 481 GCGGGACCCCACTACCTCTCTCCCGGGCCGACCCCTCGGCCCAAGCCACCTGCC 540
 QY 190 GlyThrLeuAlaAlaSerGluGlyArgTrrPlySerMetArgLysSerProLeuGly 209
 DB 541 GGGACCTGGGGGGCGGACGAGAGAGTAAAGAGAGAGCCCTCGGGGCT 600
 QY 210 GlyGlyGlySerGlyAlaSerSerGlnAlaAlaGlySerGlnIleLeuLeu 228
 DB 601 GCGCGCGCTCGGAGGCTCCAGTCAGGCGCGCTCCCAAGACAGATCTCTGCTG 657
 RESULT 8
 B0178560
 LOCUS 695 bp mRNA linear EST 30-APR-2002
 DEFINITION U1-M-EV0-bws-d-05-0-UI.r1 NIH-BMAP_EV0 Mus musculus cDNA clone
 IMAGE:5701204 5', mRNA sequence.
 ACCESSION B0178560
 VERSION B0178560.1 GI:20354052
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 695)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Dr. James Lhn, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILMIL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: PYX-5.
 FEATURES
 source Location/Qualifiers
 1..695
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5701204"
 /clone_1lb="NIH-BMAP_EV0"
 /tissue_type="whole brain"
 /dev_stage="embryo 15.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /note="Organ: brain; Vector: pYX-asc; Site_1: EcoR I;
 Site_2: Not I; The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is GGGCGTGGAA. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). Gene Discovery in the Developing Mouse Nervous
 System", supported by National Institutes of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 98 a 264 c 244 g 88 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,11e-58 Length: 695
 Score: 1104.00 Matches: 209
 Percent Similarity: 97.25% Conservative: 3
 Best Local Similarity: 95.87% Mismatches: 4
 Query Match: 34.10% Indels: 2
 DB: 14 Gaps: 2
 US-10-054-935-2 (1-614) x B0178560 (1-695)
 QY 1 MethrMetArgSerAlaValAlaPheLysAlaAlaProAlaGlyLysProGlu 20
 DB 41 ATGACCATGACATCCGCGAGTTCACAGCGGCGCCCTCGCGGCGCAACCCGAG 100
 QY 21 GlnArgLeuAspTrrGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
 DB 101 CAGGAGCTGGAGCTAGAGCGGGCTGGCGGGCGCGGAGAGAGTCCGGGGCG 160
 QY 41 AlaGluAlaHisPheLeuProArgHisArgLysLeuGlyLysGluProGlyProProLeuAla 60
 DB 161 GCGGAGGCGCATTTCTCCCGGCAATCTCAAGAGAGCGGGGCGCCCTCGGCC 220
 QY 61 SerSerGlnGlyLysSerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79
 DB 221 TCTTCCAGGGCGGGAGCGCCCTCGCTCTCAGCGGCTCGCGCGGCGCAAGGCGCG 280
 QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnSerTrrPglyLys 99
 DB 281 GCGTTGTACTCCAGCGGGGGGGGGCGGGCGGAGAGAGAGAGTGGGGCGGTTCG 340
 QY 100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyIleGlyGlyGluProAla 119
 DB 341 GTGCCCTTGCCCTTCGCGCCCGGCACTAAAGCGGCATCGCGGGGAGGCACTC 400
 QY 120 AlaAlaGlyAlaGlyCysSerProArgProLysTrrGlnAlaValLeuProIleGlnThr 139
 DB 401 GCAGCGGGCGGTGGGTGGAGCCCGGGCCAAAGTACAGGGGTGCTCCCATTCACAGCG 460
 QY 140 GlySerLeu---ValAlaAlaAlaLysGluProThrProTrrPalaglyAspLysGly 158
 DB 461 GGCCTATACGTGGTGGGGGGGGCCAAAGAGCTACGCCCTGGCGTGGGAGACAGGTGG 520
 QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro 178
 DB 521 GCGGCTCCCGCAGCTGCACCGCCTCGGACCGGGGACCCCACTACCTCTGCC 580
 QY 179 GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyArg 198
 DB 581 GGGCCACACCCCTCGGCGCCACCGCACTCTGGANCCCTGGGGCCAGTGAAGGCCAGA 640
 QY 199 TrrPlySerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSer 216
 DB 641 TGAAGAGATATTAAGAGAGAGCCCTCGGGGTGGCGGGCTCGGGAGAGCTCC 694
 RESULT 9
 B0257208
 LOCUS 686 bp mRNA linear EST 06-MAY-2002
 DEFINITION N1SC_Kol1e07.g3 Baker mouse embryo e6.5 Mus musculus cDNA clone
 IMAGE:5407980, mRNA sequence.
 ACCESSION B0257208
 VERSION B0257208.1 GI:20457961
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 686)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 REFERENCE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

**JOURNAL
COMMENT**

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
CDNA Library Preparation: J. Baker (Stanford University)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NTSC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
info@image.lnl.gov
MGI:1844716
Plate: L1AM12040 row: J column: 13
Seq primer: Sp6 primer.

BASE COUNT	198 a	179 c	193 g	116 t
ORIGIN				

Alignment Scores:	
Pred. No.:	1.6e-57
Score:	1097.00
Percent Similarity:	96.89%
Best Local Similarity:	94.67%
Query Match:	33.88%
AB:	14
	Gaps:
	length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	Gaps:

US-10-054-935-2 (1-614) x BQ257208 (1-686)

OY	178	ProG1ProProbleuAlaProthrlaPrlaPrlaaglylthrlleuAlaSerG1ngly	197
Db	12	CCGGGCCACACACCCCTCGGCCACACGCCACTCTGGGACCTCGCGGACGTAGAGGC	71
OY	198	ArgTPrLysSerMetArgLysSerProLeuEnglyGlyGlySerGlyAlaSerSer	217
Db	72	AGATGGAAAGATTAAGAGAGAGCCCTCTCGGGGTGGCGCGCTCGAGAGCTCCAGT	133
OY	218	GlnAlaAlaCylLeuLysGlnIleLeuLeuEnglnleuAspLeuIleGlnGlnGln	237
Db	132	CAGCGCCGCTGCTCAACAGATCCCTTCGCTGCATTTGACCTATCGAACACACACAG	191
OY	238	GlnGlnLeuGlnAlaLysGlnLysGlnIleGlnGlnLeuLysSerSerGluArgSPThrLeu	257
Db	192	CACAGCTGCAGGCCAAGAGAGAGAGATAGAGAGAGCTGGAAGTCCGAGAGAGATACCTC	251
OY	258	LeuAlaArgIleGluArgMetLuarArgMetGlnLeuValLysLysAspAsnGlnLys	277
Db	252	CTTGCCTGGATTGAACGTATGGAAAAGCGGAGTGCAGCTGTGGAAAGAGGATTAACGAGAC	311
OY	278	GluArgHisLysLeuPheGlnGlyArgGluThrGluArgGluGluGluGluGlnLeuSer	297
Db	312	GAAAGGCACAACTGCTCCAGGCGTATGAACTCTGAAGAGAGAGAGAGACAGAGATTGCT	371
OY	298	GlnLysIleLysLeuGlnLysGlnProGlnLeuSerGlnThrSerGlnThrLeuProPro	317
Db	372	GAGAAATATTAATTGGAGCGCCAGCGAGCTTTCGAGACCTCCACAGGCTCTGCTTCC	431
OY	318	LysProPheSerCysGlyArgSerSerLysGlnHisLysArgLysSerProPheGlySer	337
Db	432	AAGCTTTTCTCATGTGGCCGAGGTGGAAAGGAGACCAAAAGGAAACCCCATTTGGGAAT	491

Qy	338	ThrlguArqLysThrProValLysLeuAlaProGluPheSerLysValLysThrLys	357
Db	492	ACAAGAAAGAAAGACTCCTGTTAAAGAGCTGGCTCTGATTTTCAAAAGTCAAAACAAA	551
Qy	358	ThrProLysHisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCys	377
Db	552	ACTCCTTAAGCACTCTCCCATTAAGAGAGAACCCGTGTGTTCCATATACGAAACGTGTTGT	611
Qy	378	LysArqGluLeuAlaGlySerGluGluThrProGluLysProArgSerSerValAspThrPro	397
Db	612	AAAGGTGATTTGAGAGAGCAAGAAAACCCCAAGAAAAGCCCGGTCTTCAGTGATACCCCA	671
Qy	398	ProArgLeuSerThr	402
Db	672	CCAAGACTCTCGACT	686
RESULT 10			
BO881177			
LOCUS	BO881177	891 bp	mRNA
DEFINITION	AGNCNCU001_8065588 NIH_MGC_102 Homo sapiens CDNA IMAGE:6212597	linear	EST 16-AUG-2002
ACCESSION	BO881177	5', mRNA sequence.	
VERSION	BO881177.1	GI:22273185	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 891)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCM2376 row: p column: 06
High quality sequence stop: 601.

FEATURES	
source	Location/Qualifiers
1..891	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:6212597"	
/clone_1id="NH_MGC_102"	
/tissue_type="epidermoid carcinoma, cell line"	
/lab_host="DH10B (phage-resistant)"	
/note="Organ: salivary gland; Vector: pORF7; site_1: XhoI, site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NH_MGC library."	
BASE COUNT	288 a 209 c 205 g 188 t 1 others
ORIGIN	
Alignment Scores:	
Pred. No.:	3,24e-56
Score:	1077.50
Percent Similarity:	94.20%
Best Local Similarity:	93.30%
Query Match:	33.28%
DB:	14 Gaps: 2

US-10-054-935-2 (1-614) x BQ881177 (1-891)

QY 255 AsprThrLeuLeuAlaArgIleGluIuArgMetGluArgMetGluLeuValLysLysAsp 274
 Db 2 GAGAGCTCTGCGCGATGTAAGTGAAGGCGATGCGTGAAGAAGAT 61
 QY 275 AsngLysGluArgHisLysLeuPheGlnGlyTyrGluThrGluGluArgGluThr 294
 Db 62 AACGAGAAAGAAAGCAGCAAGCTGTTTCAGGCGTATGAAGTGAAGAGAGAGAAACA 121
 QY 295 GluLeuSerGluLysIleLysLeuGluGlyGlnProGluLeuSerGluThrSerGlnThr 314
 Db 122 GAGCTATCTGAGAAATTAAGTGAAGTCCAGCGAGCTTTCGAGACATCCGAGCT 181
 QY 315 LeuProProLysProPheSerCysGlyArgSerGlyLysGlyHisLysArgLysSerPro 334
 Db 182 CCGCCCTCCCAAGCCCTCTCATGTGGCGGAGTGGAAAGGACATATAAGGAATCCCA 241
 QY 335 PheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLysVal 354
 Db 242 TTTGGAAAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGGCTCGAATTTCAAAAGTC 301
 QY 355 LysThrLysThrProLysHisSerProIleLysGluGluProCysGlySerLeuSerGlu 374
 Db 302 AAACAAAAACTCTCAAGCACTCTCTATTAAAGAGAACCTGTGGTTCCTATCTGA 361
 QY 375 ThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerVal 394
 Db 362 ACTGTTTGAACGTAAGTGAAGGAGCCAGAAACCCAGAAAGCCCGCTTCTCAGTG 421
 QY 395 AsprThrProProArgLysSerThrProGlnLysGlyProSerThrHisProLysGluLys 414
 Db 422 GACACCCCAAGCAAGCTCTCCACCTCCCAAAAGGAGCCAGCACCATCCCAAGAGAA 481
 QY 415 AlaPheSerSerGluIleGluAspLeuProTyrIleSerThrThrGluMetLysLys 434
 Db 482 GCCTTCTCAGTGAAGTGAAGATTTGCCGTACCTTCCACCAAGAAATGATTTGTGT 541
 QY 435 ArgTyrPheLysGlnProProSerProLeuProLeuArgLysSerProLysLysGlu 454
 Db 542 CGTGGCAACCAAGCTCCCATCAGCGTACCATTAAGGGAATCCCTCTCCAAAGAGAG 601
 QY 455 GluThrValAlaArgCysLeuMetProSerSerValAlaGluThrSerValLeuAla 474
 Db 602 GAGACTGTAGCA-----AGTAAGCATAGAGAAAGACTTCTCTT--- 640
 QY 475 ValProSerTyr 478
 Db 641 ATACCTACTGTG 652
 RESULT 11
 AM965141 686 bp mRNA linear EST 01-JUN-2000
 LOCUS EST377214 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
 DEFINITION AM965141
 ACCESSION AM965141 GI:8154977
 VERSION AM965141.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 686)
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
 I, E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
 Quackenbush, J.
 TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 210

Seq primer: Reverse.
 FEATURES Location/Qualifiers
 source 1. 686
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_11b="MAGE resequences, MAGI"
 /note="Vector: pBluescriptsm"
 BASE COUNT 229 a 165 c 164 g 128 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,74e-56 Length: 686
 Score: 1074.50 Matches: 215
 Percent Similarity: 95.228 Conservative: 4
 Best Local Similarity: 93.488 Mismatches: 7
 Query Match: 33.188 Indels: 4
 DB: 10 Gaps: 1
 US-10-054-935-2 (1-614) x AM965141 (1-686)
 QY 229 GlnLeuAspLeuIleGluGlnGlnGlnGlnLeuAlaLysGluLysGluIleGlu 248
 Db 3 CAATTGGACCTCATCGAAGCAGCAGCAGCGCTGCGAGCCAGAGAAAGAGATCGAG 62
 QY 249 GluLeuLysSerGluArgAsprThrLeuLeuAlaArgIleGluArgMetGluArgArgMet 268
 Db 63 GAGCTGAAGTCAAGAGAGACACGCTCTCGTGGATTGAACGTAATGAAAGCGGATG 122
 QY 269 GlnLeuValLysLysAspAsngLysGluArgHisLysLeuPheGlnGlyTyrGluThr 288
 Db 123 CAGCTGGTAAAGAGATGAAGAGAAAGAAAGCAAGCTGTTTCAGGGCTATGAACT 182
 QY 289 GluGluArgGluGluThrLysLeuSerGluLysLysLeuGlnCysGlnProGluLeu 308
 Db 183 GAGAGAGAGAGAGAAACACATCTATCTGAGAAATTAAGTGAAGTGCAGCGGAGCTT 242
 QY 309 SerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGly 328
 Db 243 TCCGAGACATCCCAAGCTGTGCTCCCAAGCCCTTCTCATGTGGGAGAGTGAAGAGA 302
 QY 329 HisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAla 348
 Db 303 CATAAAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGACCTGTTAAAAAGCTGCT 362
 QY 349 ProGluPheSerLysValLysThrLysThrProLysHisSerProIleLysGluGluPro 368
 Db 363 CCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGCACTCTCTATTAAAGAGAAAGCC 422
 QY 369 CysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGlu 388
 Db 423 TGTGTTCTTATCTGAAGCTGTTGTAACGTGAATTGAGAGCCAGAAAGCCAGAGA 482
 QY 389 LysProArgSerSerValAsprThrProProArgLysSerThrProGlnLysGlyProSer 408
 Db 483 AAGCCCGGCTTCAAGTGAAGCAGCCAGCAAGCACTCTCAACCCCAAAAGGAGAGCCAGC 542
 QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrIleSerThr 428
 Db 543 ACCCATCCCAAGAGAGAAAGCTTCTCAATGAGATGAAGAAATGGCGATCTTTCACAC 602
 QY 429 Thr-GluMetTyrLeuCysArgTyrPheLysGln-ProProPro-SerProLeuProLeuArg 447
 Db 603 ACAGAGAAAGTATTTGTGTGTGTGGCAGCAACGCTCCCAATAAACGTTCCCAATTACG 662
 QY 448 ---GluSerSerProLysLysGlu 454
 Db 663 GGAATAATCTTCCAAAGAGAA 686
 RESULT 12
 BQ187281 723 bp mRNA linear EST 30-APR-2002
 LOCUS BQ187281
 DEFINITION UI-E-EJ1-ajz-b-15-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
 UI-E-EJ1-ajz-b-15-0-UI 5', mRNA sequence.

ACCESSION B0187281
VERSION B0187281.1 GI:20362832
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 723)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msosares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 REVERSE.

FEATURES
source

1..723
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-a3-b-15-0-01"
/clone_1lb="UI-E-EJ1"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-E-EJ1 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pRT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAGCA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT;
optic nerve, CCATTAGTGC; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCCA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 231 a 171 c 170 g 148 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4,88e-56 Length: 723
Score: 1073.00 Matches: 208
Percent Similarity: 99.05% Conservative: 0
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 33.14% Indels: 1
DB: 14 Gaps: 0

US-10-054-935-2 (1-614) x B0187281 (1-723)

QY 249 Glnleuylsersergluargasprthrleuleualargllleqluargmetgluargmet 268
|||||
Db 2 GAGCTGAGACGACGAGACGACGCTCTGCTCGGATTACACGTAA-GAAAGCGGATG 60
QY 269 GlnleuValIylsypasngluLysgluArgHisLysLeuphegInglYrGluThr 288

|||||
Db 61 CAGCTGTAAGAAGATGATACGAGAAAGAACGACAAAGCTGTTCCAGGCGTATGAAACT 120
QY 289 GlnGluArgGluGluThrGluLeuSerGluLysIleLysLeuGluLysGlnProGluLeu 308
|||||
Db 121 GAAGAGAGAGAGAGAGAGAGAGCTATCTGAGAAAAATTAACTGGAGTCCAGCGGAGCTT 180
QY 309 SerGluThrSerGlnThrLeuProProLysProPheSerCysGluArgSerGlyLysGly 328
|||||
Db 181 TCGAGACATCCAGACTGCGCTCCCAACCCCTTCATGCTGGCGGAGTGAAGGGA 240
QY 329 HisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAla 348
|||||
Db 241 CATTAAGAGAAATCCCATTTGGAGATACAGAAAGAAAGACTCTGTTAAAAAGCTGGCT 300
QY 349 ProGluPheSerLysValLysThrLysThrProLysHisSerProIleLysGluGluPro 368
|||||
Db 301 CCTGAATTTTCANAAAGTCAAAACAAACAAACTCTTAAGCACTCTCTATTAAAGAGAACCC 360
QY 369 CysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGlu 388
|||||
Db 361 TGTGTTCTTATCTGAACCTGTTTGAACGTGAATTGAGAGCCAAAGAACCCAGAA 420
QY 389 LysProArgSerSerValAspThrProProArgLysSerThrProGlnLysGlyProSer 408
|||||
Db 421 AAGCCCGGCTTCAGTGACACACCCACCAAGACTCCACTCCCAAAAGGAGACCCAGC 480
QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluLysPheProLysSerThr 428
|||||
Db 481 ACCCATCCAGAGAGAAAGCCCTTCTCAAGTAGATGAACATTTGCCGTACCTTCCACC 540
QY 429 ThrGluMetLysLeuCysArgTrpHisGlnProProProSerProLeuProLeuArgGlu 448
|||||
Db 541 ACAGAAATGATTTGTGCTGCTGTGGACACCACTCCCAACACGCTTACACATTACGGGA 600
QY 449 SerSerProLysLysGluGluThrValAla 458
|||||
Db 601 TCCTCTCCANAGAGAGGAGACTGTAGCA 630
RESULT 13
LOCUS B1154331 811 bp mRNA linear EST 05-JUL-2001
DEFINITION 602905281F1 NCI_CGAP_Mam3 Mus musculus cDNA IMAGE:5034548 5',
B1154331
ACCESSION B1154331
VERSION B1154331.1 GI:14614332
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 811)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lohr Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: L1AM1096 row: b column: 21
High quality sequence stop: 689.
Location/Qualifiers
1..811
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone_IMAGE="5034548"
/clone_1lb="NCI_CGAP_Mam3"

	/tissue_type="tumor, gross tissue"
	/dev_stage="10 months"
	/lab_host="DH10B"
	/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
BASE COUNT	188 a 245 c 253 g 125 t
ORIGIN	
Alignment Scores:	
Pred. No.:	8.99e-56 Length: 811
Score:	1069.50 Matches: 237
Percent Similarity:	89.21% Conservative: 11
Best Local Similarity:	85.25% Mismatches: 18
Query Match:	33.03% Indels: 13
DB:	Gaps: 1
US-10-054-935-2 (1-614) x BII54331 (1-811)	
OY	98 GlySerValProLeuProCysProProAlaThrLysGlnAlaGlyIleGlyGlyGlu 117
Dd	1 GGTTGGCGCCCTTCCTGCCCTGTGCAGCCCCCACTACTAACAACGCCGGCATCGCGGGAG 60
OY	118 ProAlaAlaAlaGlyAlaGlyCysSerProArProLysTYRGlNAAlaValLeuproile 137
Dd	61 CCAGCGCAGCGCGCGCTGTGGTGAGCCCCCGGCCCAAGATACAGCGGGTGTGCCATT 120
OY	138 GlnThrGlySerLeu--ValAlaAlaAlaLysGluProThrProtPalaeLysPlys 156
Dd	121 CAGACGGGCTTCATCTGCTGGCGGGCGGCCAAAGAAGCCATCACCCCTGGGCTGGGGAAGA 180
OY	157 GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeupro 176
Dd	181 GGTGGGGGGCTCCCCCAGCTGCCACCCTGTGGACCGGCGGACCCCCCACCACCTACT 240
OY	177 LeuProGlyProProProLeuAlaLabroThrAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
Dd	241 CTGCCCCGGGCACCAACCCCTGGCGGCCAACCGGCACCTGGTGGGACCTGGCGGCAGTAG 300
OY	197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlyGlyGlySerLysAlaSer 216
Dd	301 GGCAGATGGAAAGATATAAGGAAGGCCCTCTCGGGGGGGGGGGCGGAGCGAGCTCC 360
OY	217 SerGlnAlaAlaCysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGluGlnGln 236
Dd	361 AGTCAGGCGCCCTGCCCTCAACAGATCCCTTCTGCTGCAATTGAGCCTCATCCACAGCAG 420
OY	237 GlnGlnGlnLeuGlnAlaLysGluLysGluIleGluGlnLeuLysSerGlu-ArGAspTh 256
Dd	421 CAGCAGCAGTTGCAGGCCACAGGAGAGAGATAGAGAGCTGAAGTCCAGCAGAGATAC 480
OY	256 IleuLeuAlaArgIleGluArgMet-GluArgArgMetGlnLeuValLysLysAspAsnG 276
Dd	481 GCTCCTTCTCGGATTTGAACGTATGACAACGGCGGATGCACTGGTGGAAGAGAGATTAACG 540
OY	276 IuLysGluAlaGHisLysLeuPhneGlnIleTYrGluThGlu-GluArgGluGluLuhTrgLu 295
Dd	541 AGAAGGCANAGCCACAGCTGCTCCAGGGGCTATGAACCTGACAGAGAGAGAAAGAACCCGAG 600
OY	296 LeuSerGluLys-IleLysLeuGlnCysGlnPro-GluLeu-SerGluThrSerGlnThr 314
Dd	601 TTGTCTGAGAAACATTCAATTGAGACCGCCAGCGGGAGCTTTGCAAGACACTCCACAGCT 660
OY	315 LeuPro-ProLysProPheserCysGlyArg-SerGly-LysGlyHisLysArgLysSer 333
Dd	661 CTGCTTCGGAAGCCTTACTCATGTGGCGGGAAGTGAACATGGGACCAAGAGGACAAG 720
OY	334 ProPhieGlyserThrGluArgLysThrProValLys--LysLeuAlaProGluPheserL 353
Dd	721 CCCATTGGAAATACGAAAGAAGACTCTGTTTAAAGAAGAGTGGCTCTCGAATTTTTTCAA 780

QY	353	ysVallysthrlyserProlysthrSerPro	363	
Db	781	CAGTCACACAAAC-AGTCTTAAGCATTTCCA	811	
RESULT 14		BG421221	1087 bp	mRNA
LOCUS		BG421221		linear
DEFINITION		60245171221 NIH_MGC_14 Homo sapiens CDNA clone IMAGE:4590002 5',		EST 14-MAR-2001
ACCESSION		mRNA sequence.		
VERSION		BG421221		
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE		NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT		Unpublished (1999)		
		Contact: Robert Strusberg, Ph.D.		
		Email: cgarbs-remail.nih.gov		
		Tissue Procurement: DCTD/DTF		
		CDNA Library Preparation: Ling Hong/Rubin Laboratory		
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
		DNA Sequencing by: Incyte Genomics, Inc.		
		Clone distribution: MGC clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LLNL at:		
		http://image.llnl.gov		
		Plate: L10M1324 row: h column: 03		
		High quality sequence stop: 583.		
FEATURES		Location/Qualifiers		
SOURCE		1..1087		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:4590002"		
		/clone_1ib="NIH_MGC_14"		
		/tissue_type="renal cell adenocarcinoma"		
		/lab_host="DH10B (phage-resistant)"		
		/note="Organ: Kidney; Vector: pOTB7; Site_1: XhoI; Site_2:		
		ECORI; CDNA made by oligo-dT priming. Directionally		
		cloned into ECORI/XhoI sites using the following 5'		
		adaptor: GGCACGAC(G). Size-selected >500bp for average		
		insert size 1.8kb. Library constructed by Ling Hong in		
		the laboratory of Gerald M. Rubin (University of		
		California, Berkeley) using ZAP-cDNA synthesis kit		
		(Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	293 a	309 c	345 g	140 t
ORIGIN				
Alignment Scores:				
Pred. No.:	4.97e-55	Length:	1087	
Score:	1059.50	Matches:	240	
Percent Similarity:	81.13%	Conservative:	18	
Best Local Similarity:	75.47%	Mismatches:	44	
Query Match:	32.72%	Indels:	19	
DB:	12	Gaps:	3	
US-10-054-935-2 (1-614) x BG421221 (1-1087)				
QY	107	Proalathrllysglnalaglyrllelylglygluproalaalalaglyalaaglycysser	126	
Db	2	CCGGCACCACAGCAGCCGGGCAATTGGGGGGAGCCTGCGGACGCGGAGCGGCGTCA	61	
QY	127	ProarqProllySTYrGlnalalalaleuProilleglnthrglySerleuValAlalalala	146	
Db	62	CCCCGCCCCAAGTAACAGGCGGTGCTGCCCATTCAGACGGGCTCTCTGTCGGCGGCGC	121	
QY	147	LysgluprcothlnprotrlpalaglyasprrlysglygylalalaserProalalathrlala	166	
Db	122	AAAGGCTCTACGCCCTGTGGGCTGGGGACAAAGGTGGGGGCGGCTCCCGCGCTCCACGCC	181	

Tue Jul 15 10:16:09 2003

us-10-054-935-2.rst

Page 15

Job time : 2357 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 5650 Seconds
(without alignments)
12532.154 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372

Sequence: 1 cagctcctgacccccgcac.....aaaaaaaaaaaaaaaaaa 4372

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estcov:*
6: em_estpl:*
7: em_estpro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1950.4	44.6	3161	11	AK014463 Mus muscu
2	980.8	22.4	1304	11	AK015496 Mus muscu
3	897	20.5	1096	13	BM547318 BM547318 AGENCOURT
4	770.8	17.6	817	13	B1917709 B1917709 603183689
5	736.2	17.3	765	14	B0014717 UI-H-ED1-B0014717
6	745.8	17.1	1032	13	BM475963 AGENCOURT

Amir Spaul

7	744.6	17.0	891	14	B0898842	B0898842 AGENCOURT
8	737.4	16.9	931	14	B0948768	B0948768 AGENCOURT
9	725.6	16.6	870	14	B0229745	B0229745 AGENCOURT
10	714	16.3	715	14	BM981391	BM981391 UI-CF-FN1
11	711	16.3	729	14	B0004676	B0004676 UI-H-E10-
12	685	15.7	697	13	B1820795	B1820795 603034002
13	679.8	15.5	864	12	B6289135	B6289135 602387523
14	678.2	15.5	745	12	BE907276	BE907276 601500975
15	663.2	15.2	933	14	B0876347	B0876347 AGENCOURT
16	658.4	15.1	758	12	BE792962	BE792962 601585622
17	654.4	15.0	951	12	BE178033	BE178033 602327595
18	648.6	14.8	854	13	B1665707	B1665707 603287711
19	648.6	14.8	950	12	B6386583	B6386583 602454831
20	648.2	14.8	686	14	BM970469	BM970469 UI-CF-EC1
21	638.4	14.6	640	12	BE896063	BE896063 601438862
22	637.4	14.6	686	13	B1914733	B1914733 603184105
23	625.8	14.3	1087	12	B6421221	B6421221 602451712
24	621	14.2	723	14	B0187281	B0187281 UI-E-EJ1-
25	619	14.2	627	14	B0183599	B0183599 UI-H-EUO-
26	618.8	14.2	686	14	B0257208	B0257208 NISC_K011
27	615.4	14.1	686	10	AM965141	AM965141 EST377214
28	613.4	14.0	891	14	B0881177	B0881177 AGENCOURT
29	613.2	14.0	889	12	B6506834	B6506834 601861447
30	605.2	13.8	710	10	AV713546	AV713546 AV713546
31	604.4	13.8	617	14	BM770559	BM770559 K-EST0054
32	604.4	13.8	695	14	B0178560	B0178560 UI-M-EVO-
33	597.6	13.7	637	10	BE693616	BE693616 MR4-BT035
34	597.6	13.7	647	12	B6537537	B6537537 602565862
35	597.6	13.7	698	13	B1905419	B1905419 603167481
36	595.8	13.6	631	10	BE693636	BE693636 MR4-BT035
37	595.6	13.6	705	10	BE277857	BE277857 601120091
38	595	13.6	629	10	AV722833	AV722833 AV722833
39	594.4	13.6	608	14	B0636732	B0636732 h013006_Y
40	591	13.5	677	12	B6180581	B6180581 602329333
41	590.6	13.5	662	12	BE284375	BE284375 EST448966
42	587.2	13.4	623	13	B1457952	B1457952 603198945
43	585.8	13.4	770	12	B0106391	B0106391 602289763
44	573.2	13.1	621	13	B1032095	B1032095 PM4-MT042
45	573	13.1	585	13	B1019744	B1019744 IL3-MT026

ALIGNMENTS

RESULT 1	AK014463	3161 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus	16 days embryo head cDNA	RIKEN full-length enriched library, clone:4121402D02:PRICHEXTENSN domain containing protein, full insert sequence.		
DEFINITION	AK014463				
ACCESSION	AK014463.1	GI:12852332			
VERSION	AK014463.1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) 16 days embryo head cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493774				
PUBMED	11042159				


```
|||||
Db 719 TCGTCCCATTCAGACGGGCTATCTGTGGGGGGGCGCCAAAGACCTACGCCCTGGG 778
QY 535 CTGGGAGCAAGGTTGGGGGGGCTCCCCCTGCCACCGCTGGGACCCGGGAGCCCC 594
Db 779 CTGGGGCAAGGGTTGGGGGGGCTCCCGACCTGCCACCGCTGGGACCCGGGAGCCCC 838
QY 595 CACCACTACCTTGCCCGGGGGCGGACCCCTCGGGGCCACCGGACCGGGGAGCCCTGG 654
Db 839 CACCACCTACTTGCCCGGGGGCGGACCCCTCGGGGCCACCGGACCTGCTGGACCTTGG 898
QY 655 CGGCGACGAGGGGAGATGGAAGATGAGGAAGAGCCCTCGGGGGGTGGTGGCGCT 714
Db 899 CGGCGAGTGAAGGAGATGGAAGATGGAAGAGCCCTCGGGGGGTGGGCGGCGCT 958
QY 715 CGGGAGGCTCCATGTCAGGCGGCTGCTCAAAACAGATCTTGTCTGCAATGGACCTCA 774
Db 959 CGGGAGGCTCCATGTCAGGCGGCTGCTCAAAACAGATCTTGTCTGCAATGGACCTCA 1018
QY 775 TCGAACAGCAGCAGCAGCAGCTGAGGCGCAAGGAAAAGAGATGAGAGCTGAAATCAG 834
Db 1019 TCGAACAGCAGCAGCAGCAGCTGAGGCGCAAGGAAAAGAGATGAGAGAGCTGAAATCAG 1078
QY 835 AGAGAGCAGCAGCTCCTTGTCTGGAATGAACTATGGAAGCGGATGCACTGGTAAAGA 894
Db 1079 AGAGAGTACGCTCCTTGTCTGGAATGAACTATGGAAGCGGATGCACTGGTAAAGA 1138
QY 895 AGATATACGAGAAAGAAAGGACACAGCTGTTTCAGGGCTATGAAACTGGAAGAGAAAG 954
Db 1139 GGGATTAACGAGAAAGAAAGGACACAGCTGTTTCAGGGCTATGAAACTGGAAGAGAAAG 1198
QY 955 AAACAGAGCTATCTGAGAAATTAATTAATGAGTCCGAGCGGACCTTCCGAGCATCCC 1014
Db 1199 AAGCAGAGTGTCTGAGAAATTAATTAATGAGTCCGAGCGGACCTTCCGAGCATCCC 1258
QY 1015 AGACTGCTCCAGAGCCCTTCTCATATGTTGGGGGAGTGAAGGACATTAAGAAAT 1074
Db 1259 AGGCTGCTCCAGAGCCCTTCTCATATGTTGGGGGAGTGAAGGACATTAAGAAAT 1318
QY 1075 CCCCATTTGGAAGTACGAAAGAAAGAAAGCTGTTTAAAGAGTGGCTCCGATTTTCAA 1134
Db 1319 CCCCATTTGGAAGTACGAAAGAAAGAAAGCTGTTTAAAGAGTGGCTCCGATTTTCAA 1378
QY 1135 AAGTCAAAACAAAACCTCTAGCAGCTCTTATTAAGAGAAACCTGTGGTCTTAT 1194
Db 1379 AAGTCAAAACAAAACCTCTAGCAGCTCTTATTAAGAGAAACCTGTGGTCTTAT 1438
QY 1195 CTGAAACTGTTTGAAGTGAATGGAAGCCAAAGAACCCCGAGGAGCCCGTCTT 1254
Db 1439 CAGAAACTGTTTGAAGTGAATGGAAGCCAAAGAACCCCGAGGAGCCCGTCTT 1498
QY 1255 CAGTGGACACCCCAACCAAGATCTCCACTCCCAAAAGGAGACCCAGACCCATCCCAAG 1314
Db 1499 CAGTGGATCCCAACCAAGATCTCCACTCCCAAAAGGAGACCCAGACCCCAAG 1558
QY 1315 AGAAAGCCTTCTCAAGTGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1374
Db 1559 AGAAAGCCTTCTCAAGTGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1618
QY 1375 TGTGTGCTGGACACGCTCCCATCACCCTTACATTAAGGGAATCTCTCCAAAGA 1434
Db 1619 TGTGTGCTGGACACGCTCCCATCACCCTTACATTAAGGGAATCTCTCCAAAGA 1678
QY 1435 AGGAGGAGACTGTAGCAAGGTGTGTATGATCAATGTTGCGAGGAAACCTCACTCT 1494
Db 1679 AGGAGGAGACTGTAGCAAGGTGTGTATGATCAATGTTGCGAGGAAACCTCACTCT 1738
QY 1495 TGGCTGTCTCTTGTGAGGAGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCA 1554
Db 1739 TGGCTGTCTCTTGTGAGGAGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCA 1798
QY 1555 ACCTTTGGAAACCTGATGACAGTGTGTTTGAAGCGGATGCAAAACTGGAGCTGG 1614
|||||

Db 1799 ACATTTTGGAAACCTGATGACAGTGTATTTTCAAAAGGAGCATGCAAAACTGGAGCTAG 1858
QY 1615 ATGAGAGAGAGAGAAAGATGAGATTTTCAAGAGATCAGAGAAACAAAGATTTTACAGC 1674
Db 1859 ACAGAAAGAGAGAGAAAGATGAGATTTTCAAGAGATCAGAGAAACAAAGATTTTACAGC 1918
QY 1675 GACTGAGCTCAGAGATGATTAAGAAAGAAAGATTCAGAGATCAGAGCTGAGTTCCT 1734
Db 1919 GACTGAGCTCAGAGATGATTAAGAAAGAAAGATTCAGAGATCAGAGCTGAGTTCCT 1978
QY 1735 CATTTTCCCTGAGCCAGATGATGTTGAAGTTGATGATTAACCCCTTGTGGCTGTG 1794
Db 1979 CATTTTCCCTGAGCCAGATGATGTTGAAGTTGATGATTAACCCCTTGTGGCTGTG 2038
QY 1795 TAGCATTTGAGAGCAGCTTACCAAAATTAATCTCCAGAGATTTGAGTACCTACCGGTGG 1854
Db 2039 TAGCATTTGAGAGCAGCTTACCAAAATTAATCTCCAGAGATTTGAGTACCTACCGGTGG 2098
QY 1855 ATGAGCTGAGCCGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914
Db 2099 ATGAGCGAAGCCGTTGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGAT 2158
QY 1915 GGAATATGCTGTGTGGCAAGAACCTGTCTTCAATGATGATGATGATGATGATGATGAT 1973
Db 2159 GGAATATGCTGTGTGGCGC -GATAGCCTGTCTTCAATGATGATGATGATGATGATGAT 2217
QY 1974 GAGTGGAGAGACCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2033
Db 2218 GAATAGAGAGACCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2277
QY 2034 TACCTTTTCACTTCTCTGA -CTTGTGTTTCACTTCTCTCTCTCTCTCTCTCTCTCTCT 2082
Db 2278 TGCCTTTTCACTTCTCTCTGA -CTTGTGTTTCACTTCTCTCTCTCTCTCTCTCTCTCTCT 2237
QY 2083 -AACTCTTCACTTCTCTCTGA -CTTGTGTTTCACTTCTCTCTCTCTCTCTCTCTCTCTCT 2135
Db 2238 ACAAGAAAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2394
QY 2136 TCCCTTTTGGGAAATGGGCTCTCAAGCTTAAGCTTAAGGAGGAGTTCAGAGTTT 2195
Db 2395 TCCCTTTTGGGAGATGAG -CTTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2450
QY 2196 CAGGGTCTGTCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2255
Db 2451 CAGGGGCTCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2506
QY 2256 ACATGTGAAGATGAGACGAGCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2315
Db 2507 -CGTGTCTAAGATGAGCAACCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2565
QY 2316 CCGTGTGACCCCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2375
Db 2566 CCGTGTGCGCTTCTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2604
QY 2376 ACCTCACCATGAGCTTTAGAGCCAGAAAGAGATGAGATGAGATGAGATGAGATGAGAT 2435
Db 2605 -TCACGCGAAGCTTTAGAGGCGCGAGAGAT -GCATGAGATGATGAGATGAGATGAGAT 2656
QY 2436 TCTTCAAGAGTGAAGAAAGATGAGAGATGATGATGATGATGATGATGATGATGATGAT 2495
Db 2657 TTTCTCAGAGAGGAGGAGATGAGAG -TTTAACTGAGAGGATGATGATGATGATGATGAT 2698
QY 2496 TTTCAATTTGAAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2555
Db 2699 TTTCTTTTAACTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2757
QY 2556 TTTTCACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2612
Db 2758 TTTTCACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2817
QY 2613 GAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2671
Db 2818 AGAGATTGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2877
```

RESULT 2	LOCUS	DEFINITION	AK015496	1304 bp	mRNA	linear	HTC 19-JAN-2002
AK015496			Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:430463f05:hypothetical protein, full insert sequence.				
ACCESSION			AK015496				
VERSION			AK015496.1				
KEYWORDS			HTC; CAP trapper				
SOURCE			Mus musculus (strain:J:OlaHsd/J:Ola)				

ORGANISM

REFERENCE
1. Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Meth. Enzymol.* 303, 19-44 (1999)
99279253

TITLE	Normaliza- tion and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	204939374
PUBMED	11042159
REFERENCE	3

TITLE	AUTHORS	JOURNAL	MDLINE	REFERENCE
Genome Res. 10 (11), 1757-1771 (2000)	Watanabe, T., Watanabe, M., Inoue, Y., Kita, A. and Hayashizaki, Y.	Genome Res. 10 (11), 1757-1771 (2000)	20530913	1
Genome Res. 10 (11), 1757-1771 (2000)	Kawai, J., Shinnaga, A. and Hayashizaki, Y.	Genome Res. 10 (11), 1757-1771 (2000)	11076861	4

TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (5821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5 (bases 1 to 1304)
AUTHORS	Adachi,J., Aizawa,K., Akhira,S., Arimura,H.

JOURNAL

COMMENT

FEATURES
SOURCE

CDS

Wagner,I., Mashio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsch,G., Blake,J., Boffelli,D., Bojunga,N., Carlinetti,P., de Bonaldo,M.F., Brownstein,M.J., Bulc,C., Flecher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Momberts,P., Rling,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzyki,H., Toyok-Oka,K., Wang,K., Weltz,C., Whitaker,C., Wilmig,L., Wyshaw-Borisi,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S., Hayashizaki,Y.

Functional annotation of a full-length mouse cDNA collection
Nucleotide 409 (6821), 685-690 (2001)
21085660
1121/7851

5 (bases 1 to 1304)

Aachari,J., Alizawa,K., Akahira,S., Akimura,T., Aono,H., Aral,A., Atakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bulc,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirota,T., Horii,F., Hume,D., Imochai,K., Ishii,Y., Itoh,M., Itawa,M., Kakuwasa,T., Kato,H., Kawaj,J., Koijma,T., Miyazaki,A., Nishi,K., Nomura,K., Kurihara,C., Matsuyama,T., Mizutani,K., Ono,H., Koude,M., Numazaki,K., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Salto,H., Salto,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schimml,L., Shibata,K., Shibata,Y., Shingawara,A., Shiraki,T., Tanaka,T., Teijima,Y., Toyota,T., Yamamura,T., Takahashi,F., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission
Submitted (10-JULY-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gs.c.riken.go.jp, Fax:81-45-503-9216]
<http://genome.gsc.riken.go.jp/>, Tel.81-45-503-9222,
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to Prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGGAGGAGGAGCGATCCAGAAGCTCTCTTTTCTTTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot -10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGGAGGTTCGATCGATTAAATTAATTCGCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after bulk excision from lambdaBFLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

Location/Qualifiers

1..1304

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="PANTOM.DB:4930463F05"

/db_xref="MGD:MGI:1908189"

/db_xref="taxon:10090"

/clone="4930463F05"

/sex="male"

/tissue_type="testis"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

17..1129

/note="evidence:NAS hypothetical protein putative"

/codon_start=1

/protein_id="BAB29868.1"


```

|||||
Db 181 TAGAGACTTAGAGGAGGTAGGGGAGAGTGTGGAATAGTCTCTTGGCTGCAAA 240
2837 TGTCTACATCTTGAACAACAGAGTACTATAGCTTCCATCTACTTTGTAAGA 2896
241 TGTCTACATCTTGAACAACAGAGTACTATAGCTTCCATCTACTTTGTAAGA 300
2897 TAATTGTATGTGTACATCTTGTGCTCTCCCTCCGTTTGTAAATATCAGATA 2956
301 TAATTGTATGTGTACATCTTGTGCTCTCCCTCCGTTTGTAAATATCAGATA 360
2957 GCAGTCCAGAGCCACTTGTGCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 3016
361 GCAGTCCAGAGCCACTTGTGCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 420
3017 AGAGCCAGAGCCCTGTGCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 3076
421 AGAGCCAGAGCCCTGTGCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 480
3077 TGTGTGAACAACAGTGAAGGCTCTAATCTCTAGAAGTAGGGCTTTATCTTAAGA 3136
481 TGTGTGAACAACAGTGAAGGCTCTAATCTCTAGAAGTAGGGCTTTATCTTAAGA 540
3137 GAATATGTGCCAGATTAATTAAGCACTTTAGAGAGAAAGCAAGGTATAGGTGTGTG 3196
541 GAATATGTGCCAGATTAATTAAGCACTTTAGAGAGAAAGCAAGGTATAGGTGTGTG 600
3197 GCTGGCCCATGCTGAGCAGACAGAAAGAGATGATGATGATGATGATGATGATG 3256
601 GCTGGCCCATGCTGAGCAGACAGAAAGAGATGATGATGATGATGATGATGATG 660
3257 GTTCTCAGAGGGGCTCTCCACTGCTAAAGTTTGTGAGATGTGATCTGCTCTCTG 3316
661 GTTCTCAGAGGGGCTCTCCACTGCTAAAG-TTTTGTGAGATG-TGATCTGCTCTCTG 718
3317 ATTGACTTTTAAAGAAATTAATGCGACACATGATGATGATGATGATGATGATG 3376
719 ATTGACTTTTAAAGAAATTAATGCGACACATGATGATGATGATGATGATGATG 778
3377 CTCTATTTCTCCCTTTTGTGCTGCTGCTGTG 3410
779 TCTTATTTCTCTTGTGTGTGTGTGTGTG 812

RESULT 5
B0014717/c 765 bp mRNA linear EST 26-MAR-2002
LOCUS B0014717
DEFINITION UT-H-ED1-act-1-04-0-UI-s1 NCI_CGAP_ED1 Homo sapiens cDNA clone
IMAGE:5833443 3', mRNA sequence.
ACCESSION B0014717
VERSION B0014717.1 GI:19739618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 765)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA-ies.
FEATURES
source 1..765
Location/Qualifiers

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5833443"
/clone_lib="NCI_CGAP_ED1"
/lisse_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Organ: Left Public Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site:1: EcoR I;
Site:2: Not I; NCI_CGAP_ED1 is a normalized cDNA library
containing the following tissue(s): Chondrosarcoma cell
line C55. The library was constructed according to Bonaldo
Lennon and Soares, Genome Research, 6:791-806, 1996.
First strand cDNA synthesis was primed with an oligo-dT
primer containing a Not I site. Double stranded cDNA was
ligated to an EcoR I adaptor, digested with Not I, and
cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GCTCAAGGCT.
TAG_LIB=UT-H-ED1
TAG_LTISSUB=chondrosarcoma
TAG_SEQ=CGTCAAGCT"

BASE COUNT 158 a 176 c 180 g 250 t 1 others
ORIGIN

Query Match 17 38; Score 756.2; DB 14; Length 765;
Best Local Similarity 99.58; Pred. No. 5,7e-92;
Matches 758; Conservative 0; Mismatch 4; Indels 0; Gaps 0;

946 AGAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGAGCCGAGGAGCTTCCG 1005
765 AGAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGAGCCGAGGAGCTTCCG 706
1006 AGACATCCAGACTGTGCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGGACATA 1065
705 AGACATCCAGACTGTGCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGGACATA 646
1066 AAAGAAATCCCATTTTGAAGTACAGAAAGAAACATCTCTTAAAGAGCTGCTCTG 1125
645 AAAGAAATCCCATTTTGAAGTACAGAAAGAAACATCTCTTAAAGAGCTGCTCTG 586
1126 AATTTCAAAAGTCAAAACAAACCTCTAAGCACTCTCTAATGAAGAGAACCTGTG 1185
585 AATTTCAAAAGTCAAAACAAACCTCTAAGCACTCTCTAATGAAGAGAACCTGTG 526
1186 GTTCTTAATCTGAACCTGTTTGTAAACGTGAATGAGAGCCAGAGAAAGC 1245
525 GTTCTTAATCTGAACCTGTTTGTAAACGTGAATGAGAGCCAGAGAAAGC 466
1246 CCCGCTTCACTGAGACACCCACCAAGACTCTCACTGCCCAAAAGGACCCACACC 1305
465 CCCGCTTCACTGAGACACCCACCAAGACTCTCACTGCCCAAAAGGACCCACACC 406
1306 ATCCAGAGAGAAACCTTCTCAAGTGAATGAAGATTTGCGATACCTTCCACACAG 1365
405 ATCCAGAGAGAAACCTTCTCAAGTGAATGAAGATTTGCGATACCTTCCACACAG 346
1366 AATGTATTTGTGTGTGGCAGCAGCTCCGCCATCACTTACCATTAAGGAGATCT 1425
345 AATGTATTTGTGTGTGGCAGCAGCTCCGCCATCACTTACCATTAAGGAGATCT 286
1426 CTCGAAGAGAGAGAGACTGTGACAGAGTGTCTATGCCATCAAGTGTGAGAGAAA 1485
285 CTCGAAGAGAGAGAGACTGTGACAGAGTGTCTATGCCATCAAGTGTGAGAGAAA 226
1486 CTTCAGTCTTGCTGTCTCTTGTGAGGAGACCACTGAGAGGCTCTTAAGGAGCCAA 1545
225 CTTCAGTCTTGCTGTCTCTTGTGAGGAGACCACTGAGAGGCTCTTAAGGAGCCAA 166
1546 ATCTTCAGAGCTTTTGGAGAACCTGATGACAGTGTGTTTGAAGGAGGATGCAAAAC 1605
|||||

```

Db	165	ATCCTTCAGACCTTTTGGAGAAACCTGGATGACAGTGCTTTTTCAGAGCGGATCGCAAAAC	106
QY	1606	TGGAGCTGGATGAGACAGAAAGAAAGATGGATATTACAGAGATCAGGACACAAAGAA	16655
Db	105	TGGAGCTGGATGAGAAAGAAAGAAAGATGGATATTACAGAGATCAGGACACAAAGAA	46
QY	1666	TTTATACAGCAGCTCAGCTCAGAAATGTTATTAAGAAAGAA	1707
Db	45	TTTATACAGCAGCTCAGCTCAGAAATGTTATTAAGAAAGAA	4
RESULT 6			
LOCUS	BM475963	1032 bp	mRNA
DEFINITION	AGNCNCOURT 6480838 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:5579934		EST 05-FEB-2002
ACCESSION	BM475963		
VERSION	BM475963.1		
KEYWORDS	5', mRNA sequence.		
SOURCE	EST.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 1032)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov		
	Plate: LHAM12337	row: g	column: 07
	High quality sequence stop: 653.		
FEATURES	Location/Qualifiers		
source	1..1032		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5579934"		
	/clone_id="NIH_MGC_67"		
	/tissue_type="retinoblastoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: eye; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo Average insert size 1.75 kb. Library constructed by Life Technologies."		
BASE COUNT	284 a	226 c	245 g
ORIGIN	277 t		
Query Match	17.1%;	Score 745.8;	DB 13; Length 1032;
Best Local Similarity	94.4%;	Pred. No. 1.2e-90;	
Matches	852;	Conservative 0;	Mismatches 37; Indels 14; Gaps 7;
QY	1495	TGGCTGTTCTCTTCTTGGAGGAGCAGCTAGTAGCCTCTTAAGGACCAATCTCTCAG	1554
Db	36	TAGCAATTCCTCTTCTTGGAGGAGCAGCTAGTAGAGCCCTTAAGGACCAATCTCTCAG	95
QY	1555	ACCTTTTGGAGAACCTGGATGATACAGTGTGTTTGGAAAGCGCATGCAAACTGAGCTGG	1614
Db	96	ACCTTTTGGAGAACCTGGATGATACAGTGTGTTTGGAAAGCGCATGCAAACTGAGCTGG	155
QY	1615	ATGGAAGAGAGAAAGAAAGATGGATATTACAGAGATCAGGAGCAAAAGATTTTACAGC	1674
Db	156	ATGGAAGAGAGAGAAAGAAAGATGGATATTACAGAGATCAGGAGCAAAAGATTTTACAGC	215
QY	1675	GACTGCAGCTCAGATGTATTAAGAAAGAAAGATTCAGAGATCTGAGCTGAGCTTACT	1734
Db	216	GACTGCAGCTCAGATGTATTAAGAAAGAAAGATTCAGAGATCTGAGCTGAGCTTACT	275

OY	1735	CATTTCCTCCGAGACCAGATGATGCTGGAAGTTTGATGATTTACCCCTCTTGCGCTGTG	1794
Db	276	CATTTTTCCTCCGAGACCAGATGATGATGCTGGAAGTTTGATGATTTACCCCTCTTGCGCTGTG	335
OY	1795	TAGCATTTGGACGACCACTTACCAAAATTAATCTCACAGAAATTTTGAGCTACCCCTGGTTGG	1854
Db	336	TAGCATTTGGACGACCACTTACCAAAATTAATCTCACAGAAATTTTGAGCTACCCCTGGTTGG	395
OY	1855	ATGAGCGGTAGCCGATGCGAGATTGGAGATCCAGAAAGAACAACACCTCACCGGACGTGA	1914
Db	396	ATGAGCGGTAGCCGATGCGAGATTGGAGATCCAGAAAGAACAACACCTCACCGGACGTGA	455
OY	1915	GGAATATAGCTGTGCTGGCAGAAACCCTGTCTTCAGATAGTTGATGCAATGGCATTTCCGAG	1974
Db	456	GGAATATAGCTGTGCTGGCAGAAACCCTGTCTTCAGATAGTTGATGCAATGGCATTTCCGAG	515
OY	1975	AGTGGCAGAGACCTGTATATGTGACCTTTGGCCACATATGTTATCATCGCTGATAT	2034
Db	516	AGTGGCAGAGACCTGTATATGTGACCTTTGGCCACATATGTTATCATCGCTGATAT	575
OY	2035	ACCCTTTCATACCTTCCTTGACCTTTGTTTTCATTACTCTGATTCACAAAAACCTCTTCAT	2094
Db	576	ACCCTTTCATACCTTCCTTGACCTTTGTTTTCATTACTCTGATTCACAAAAACCTCTTCAT	635
OY	2095	TCGGCTAATGTGTGAGTTATGGAGGGTGATTTGGGATTTCTTT--CCCTTTTGGGAATG	2153
Db	636	TCGGCTAATGTGTGAGTTATGGAGGGTGATTTGGGATTTCTTTTGGGGAATG	695
OY	2154	GGCTCTCAAGCTAAAGCTATAGATATGGCAGATTCAGAAAGTTTCA--GGGGTCTGTCTAT	2212
Db	696	GGCTCTCAAGCTAAAGCTATAGATATGGCAGATTCAGAAAGTTTCA--GGGGTCTGTCTAT	755
OY	2213	ACA--TTTGCCCTATGTTAAAGGGGTTAAAGGGCTCTCTTCAATTGACATGTGG--AAGATG	2269
Db	756	ACATTTTGCCCTATGTTAAAGGGGTTAAAGGGGCTCTCTTCAATTGACATGTGG--AAGATG	815
OY	2270	AAGCAGGCCCTT--CCTTTAGACGTGT--GCCTCGATGCGACTCTTTTCACCTCGTGACACC	2327
Db	816	AAGCAGGCCCTTCCCTCTTAAAGCTGTGGCCCTGCGATGACCTCTTCCCTCCCGGGTACCA	875
OY	2328	CTCCTTATAGTGGG-----TATAGATGATTTTAAACCTTAATAATAAACAACAACCTC	2380
Db	876	CCCTCCCTTATAGGGGGTAAATGTGCAATTTTAAACCTTAATAATAAACAACAACCTC	935
OY	2381	ACC 2383	
Db	936	CCC 938	
RESULT 7			
B0898842		891 bp	linear
LOCUS			EST 16-AUG-2002
DEFINITION			AGNCOCURT 8121821 Lupski_dorsal_root_ganglion Homo sapiens cDNA
ACCESSION			clone IMAGE:6181536 5', mRNA sequence.
VERSION			B0898842
KEYWORDS			B0898842.1 GI:22290856
SOURCE			EST.
ORGANISM			human.
REFERENCE			Homo sapiens
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
COMMENT			1 (bases 1 to 891)
			NIH-MGC http://mgc.ncl.nih.gov/ .
			National Institutes of Health, Mammalian Gene Collection (MGC)
			Unpublished (1999)
			Contact: Robert Strausberg, Ph.D.
			Email: cgaps-remail.nih.gov
			Tissue Procurement: Dr. James R. Lupski
			cDNA Library Preparation: Life Technologies, Inc.
			DNA Sequencing by: Agencourt Bioscience Corporation
			Clone distribution: MGC Clone distribution information can be
			found through the I.M.A.G.E Consortium/LNL at:

Db	61	GTGATGTGTGGCTATGGCTTTTCATTTGTAACCTCATTCTCTAGAGAGATGGGCTCTC	120
Qy	3466	TATATAGGGAACCTCTGTAACCTTCATTTGCGCAAGATGTAGAGAAATAGACTTAA	3523
Db	121	TATATAGGGAACCTCTGTAACCTTCATTTGCGCAAGATGTAGAGAAATAGACTTAA	180
Qy	3526	TTCCACTAGGGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAAACTGGGCAG	3585
Db	181	TTCCACTAGGGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAAACTGGGCAG	240
Qy	386	ATTTTCTTGTGTCATCATTTTAATGTCGAGGCTGTCTCAGTTTCTTACTTTACCT	3645
Db	241	ATTTTCTTGTGTCATCATTTTAATGTCGAGGCTGTCTCAGTTTCTTACTTTACCT	300
Qy	3646	ATGTGATTTTCTTCGTAACGTGTCCAAAAAGAAAAAGACCACATCGATGTCTCTTAC	3705
Db	301	ATGTGATTTTCTTCGTAACGTGTCCAAAAAGAAAAAGACCACATCGATGTCTCTTAC	360
Qy	3706	TTTGTCTTCTTGATCCCTCAGTTTCTTCTGATTTTCAGCATGTGTGGGTTCTTAATTTTG	3765
Db	361	TTTGTCTTCTTGATCCCTCAGTTTCTTCTGATTTTCAGCATGTGTGGGTTCTTAATTTTG	420
Qy	3766	GGTATGATTTAGCAAAATTTAAACATTTGTTTGTGCTTACCAGGGGACATCCCATGTT	3825
Db	421	GGTATGATTTAGCAAAATTTAAACATTTGTTTGTGCTTACCAGGGGACATCCCATGTT	480
Qy	3826	CTGACTTGAATAGACTAGAGAAGATCCAGCAGGCTCATCTGTGCCAGATTTAAGTAGAT	3885
Db	481	CTGACTTGAATAGACTAGAGAAGATCCAGCAGGCTCATCTGTGCCAGATTTAAGTAGAT	540
Qy	3886	TCTATTTTCTTGTGTTCTCCCTGCTGTCAGACCTCTTATTTATGTCCCTCTTCTAG	3945
Db	541	TCTATTTTCTTGTGTTCTCCCTGCTGTCAGACCTCTTATTTATGTCCCTCTTCTAG	600
Qy	3946	GTTAATTTCTCCTTTGATTTGACTTTGTTGTAAGAGAGAGTTGACAGTATTAACAAGT	4005
Db	601	GTTAATTTCTCCTTTGATTTGACTTTGTTGTAAGAGAGAGTTGACAGTATTAACAAGT	660
Qy	4006	TCCAAGTGCAAAATTTACAGTGTGTAGAGTGTGGGGG--AAATTAAGTCTTA--TTTTTC	4062
Db	661	TCCAAGTGCAAAATTTACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTAATTTTTTC	720
Qy	4063	CCCTACATGGG--ATPACACACG--GTGAATTCATGTTCAACG--AAGGCCCTGCAGTCTC	4119
Db	721	CCCTACATGGGAAATACACACTGGTGAATTCATCTTCACTGAAGGCCCTGCAGTCTC	780
Qy	4120	CTAAACATA---GTTGTGTTGTTTCTTTTACAAA--GTTTAAGCTAGTGTAAATTA	4173
Db	781	TTAAACCTAAAGTTGGGTTGTTTCTTTTAAACAAAAGTTTAAAGTAGGCTTAATTA	840
Qy	4174	TTAATAAAAAA---TTGCTGTCTGTAC--TTACGCTTGTGTTTATGCCATT	4223
Db	841	TTAATAAAAAAATTTGGCTTGCTGTGCTGCACTTTCAGCTTTGTGTTATGCCAATT	895
RESULT 9			
LOCUS	BQ229745	870 bp	mrna
DEFINITION	AGNCOURT_7510022 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042882	linear	EST 02-MAY-2002
ACCESSION	BQ229745		
VERSION	BQ229745.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	EMMAYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 870)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

FEATURES	source
EMAIL: c9aphb-rf@mail.nih.gov	
Tissue Procurement: ATCC	
CDNA Library Preparation: Life Technologies, Inc.	
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)	
Genomic Sequencing by: Agencourt Bioscience Corporation	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at:	
http://image.llnl.gov	
Plate: LRM13282 row: p column: 19	
High quality sequence stop: 556.	
Location/Qualifiers	
1. 870	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_image:6042882"	
/clone_idb="NIH_MGC_92"	
/tissue_type="embryonal carcinoma, cell line"	
/lab_host="DH10B (phage-resistant)"	
/note="Organ: testis; Vector: pCMV-Sport6; site_1: NotI; site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."	
BASE COUNT	171 a 289 c 302 g 107 t 1 others
ORIGIN	
Query Match	16.6%; Score 725.6; DB 14; Length 870;
Best Local Similarity	99.0%; Pred. No. 6.4e-88;
Matches	761; Conservative 0; Mismatches 5; Indels 3; Gaps 3
113	GGCCCTCCGGCGGCATCTCTGACGACGACTGACTACGACGGCGCTCGCGCGCTGGG 172
1	GGCCCTCCGGCGGCATCTCTGACGACGACTGACTACGACGGCGCGCTCGCGCGCTGGG 60
173	CGGCGCCGAGAGACGAGCCTGGGGCGCGCGAAGCCACTTCCCTCCCGCGACGTAAGCT 232
61	CGGCGCCGAGAGACGAGCCTGGGGCGCGCGAAGCCACTTCCCTCCCGCGACGTAAGCT 120
233	CAGAAGACCCGGGGCCCCCGCTGGCCTCTCCCGAGGGCGGGAGCCCGCGGCTTCCCGGGC 292
121	CAAGGAGCCGGGGCCCCCGCTGGCCTCTCCCGAGGGCGGGAGCCCGCGGCTTCCCGGGC 180
293	CGGCTGGCGGGCAAGGGCCGGGGGCTTCTTACTTCCCGGGCGGGGGCGCCCGGGCGAGCA 352
181	CGGCTGGCGGGCAAGGGCCGGGGGCTTCTTACTTCCCGGGCGGGGGCGCCCGGGCGAGCA 240
353	GGAAGAGAGCTGGGGCGGTTGGTGCTCCCTTGGCCCTGTCCCGCCCGCGCCACCAAGCAAGC 412
241	GGAAGAGAGCTGGGGCGGTTGGTGCTCCCTTGGCCCTGTCCCGCCCGCGCCACCAAGCAAGC 300
413	CGGCATTGGGGGGGAGCCTGCCGCGACCGGGAGCCGGCTGGCAAGCCCCCGGCCCAAGTATCA 472
301	CGGCATTGGGGGGGAGCCTGCCGCGACCGGGAGCCGGCTGGCAAGCCCCCGGCCCAAGTATCA 360
473	GGCGGTGTCGCATTTCAGACGGGGCTCTCGTGGCGCGCGGCCAAGAAGCTTACGCCCTG 532
361	GGCGGTGTCGCATTTCAGACGGGGCTCTCGTGGCGCGCGGCCAAGAAGCTTACGCCCTG 420
533	GGCTGGGAGCAAGGGTGGGGCGGCTCCCGCGCTGCACCGCTTCGAGCCGGCGGGAGCC 592
421	GGCTGGGAGCAAGGGTGGGGCGGCTCCCGCGCTGCACCGCTTCGAGCCGGCGGGAGCC 480
593	CCGACCACTACTCTGGCCCGGGCGCGCACCCCTTGGCGGCGACCGGACACCGCGGGAGCCCT 652
481	CCGACCACTACTCTGGCCCGGGCGCGCACCCCTTGGCGGCGACCGGACACCGCGGGAGCCCT 540
653	GGCGGCCAGCGAGGGCAGATGGAAGATATGAGGAAGAGCCCTTCGCGGGGTGTGGCGG 712
541	GGCGGCCAGCGAGGGCAGATGGAAGATATGAGGAAGAGCCCTTCGCGGGGTGTGGCGG 600
713	CTCGGAGACCTTCACGTCAGCGCGCTGCTCAAAAGATCTTCTGCTGCAATTTGGACTT 772
601	CTCGGAGACCTTCACGTCAGCGCGCTGCTCAAAAGATCTTCTGCTGCAATTTGGACTT 660

QY 773 CATGACAGCAGCAGCAGCTGTCAGAGCAAGAAAAGATCGAGCTGAAGTC 832
|||||
Db 661 CATGACAGCAGCAGCAGCAGCTGTCAGAGCAAGAAAAGATCGAGCTGAAGTC 720
QY 833 AG-AGAGAGCAGCAGCTCC-TTGCTCGGATTG-AACTATGGAAGCGCG 878
|||||
Db 721 ANAAAGAAAACGCTCTTGTGCGATTGAAACGATATGGAAGCGCG 769

RESULT 10
BM981391/c 715 bp mRNA linear EST 21-MAR-2002
LOCUS UI-CF-EN1-adh-f-11-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
DEFINITION
ACCESSION BM981391
VERSION BM981391.1 GI:19603832
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 715)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul.mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
Source
Location/Qualifiers
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-adh-f-11-0-UI"
/clone_1id="UI-CF-EN1"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pRT3-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an Ecor I adaptor, digested with Not
I, and cloned directionally into pRT3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG_LIB=UI-CF-EN1
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ=CTGCTCAGGT"

BASE COUNT 263 a 132 c 123 g 196 t 1 others
ORIGIN

Query Match 16.3%; Score 714; DB 14; Length 715;
Best Local Similarity 99.9%; Pred. No. 2.6e-86;
Matches 714; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3644 CTATGTGATATTTCTGCTAGCTGTCGCAAAAAAGAACCAATCACTGCTCTG 3703
|||||
Db 715 CTATGTGATATTTCTGCTAGCTGTCGCAAAAAAGAACCAATCACTGCTCTG 656

QY 3704 ACTTTGTTCTTTGATCCCTCAGTTCTCTTGATTTACAGATGTGCGGTTCTTAATT 3763
|||||
Db 655 ACTTWTGTTCTTTGATCCCTCAGTTCTCTTGATTTACAGATGTGCGGTTCTTAATT 596

QY 3764 TGGGTAGAGTTACCAATTAACATGTTGTTGTCCTACCAAGGGGAGCTCCAGT 3823
|||||
Db 595 TGGGTATGAGTTACCAATTAACATGTTGTTGTCCTACCAAGGGGAGCTCCAGT 536

QY 3824 TTCTGACTTGAAGTAGACTGAGAAATCCAGAGGTGCTATCTGCGCAGATTAAGTAG 3883
|||||
Db 535 TTCTGACTTGAAGTAGACTGAGAAATCCAGAGGTGCTATCTGCGCAGATTAAGTAG 476

QY 3884 ATTCATATTTCTGTTGTTCTCCCTCCTGAGACCTCTTATTTATTTGTCCTCTTCT 3943
|||||
Db 475 ATTCATATTTCTGTTGTTCTCCCTCCTGAGACCTCTTATTTATTTGTCCTCTTCT 416

QY 3944 AGTTTAATTCCTCTTGTATTTGACTTTGTGAGAAGAGGTGGACAGTATAGCAAA 4003
|||||
Db 415 AGTTTAATTCCTCTTGTATTTGACTTTGTGAGAAGAGGTGGACAGTATAGCAAA 356

QY 4004 GTTCCAAAGTGCMAATATACAGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCC 4063
|||||
Db 355 GTTCCAAAGTGCMAATATACAGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCC 296

QY 4064 CTACATGGGATACACACACTGCAATTCATCTTCACTGAAGGCCCTGCACTTCTCTAA 4123
|||||
Db 295 CTACATGGGATACACACACTGCAATTCATCTTCACTGAAGGCCCTGCACTTCTCTAA 236

QY 4124 AACATAGTGTGTTGTTTCTTAAAGAGTTAAGTAGAGTAAATTAATTAATTAATTA 4183
|||||
Db 235 AACATAGTGTGTTTCTTAAAGAGTTAAGTAGAGTAAATTAATTAATTAATTAATTA 176

QY 4184 TTGCTGTCTGCTACCTCAGCTTGTGTTTAAAGCCATTTCAATTTGTTGTTGTTGT 4243
|||||
Db 175 TTGCTGTCTGCTACCTCAGCTTGTGTTTAAAGCCATTTCAATTTGTTGTTGTTGT 116

QY 4244 AATTCATACCTTTTGATACCACTTCTGATGTGTTAAATTTGTTGTTGTTAAATCTTA 4303
|||||
Db 115 AATTCATACCTTTTGATACCACTTCTGATGTGTTAAATTTGTTGTTGTTAAATCTTA 56

QY 4304 TAAAGAGTTCAATTTGTAATTAACATATGTCGCTGTTAAAAAATTAATTAATTAATTA 4358
|||||
Db 55 TAAAGAGTTCAATTTGTAATTAACATATGTCGCTGTTAAAAAATTAATTAATTAATTA 1

RESULT 11
B0004676/c 729 bp mRNA linear EST 26-MAR-2002
DEFINITION UI-H-E10-ayp-e-01-0-UI.s1 NCI-CGAP_E10 Homo sapiens cDNA clone
LOCUS IMAGE:5841768 3', mRNA sequence.
ACCESSION B0004676
VERSION B0004676.1 GI:19729576
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 729)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov

Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 The following repetitive elements were found in this cDNA
 sequence: 1-37, >POLY_A\$Simple_repeat (matched complement)
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES

source
 1. 729
 /location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5841768"
 /clone_11b="NCI_CGAP_E10"
 /tissue-type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
 with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
 NCI_CGAP_E10 is a cDNA library containing the following
 tissue(s): Chondrosarcoma. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (drr)18 tail. The
 sequence tag for this library is ACACCTGCAC.
 TAG_L1B=UI-H-E10
 TAG_TISSUE=chondrosarcoma
 TAG_SEQ=ACACCTGCAC"
 BASE COUNT 248 a 141 c 137 g 203 t
 ORIGIN

Query Match 16.3%; Score 711; DB 14; Length 729;
 Best Local Similarity 100.0%; Pred. No. 6.4e-86;
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3476 CCGCGTGAACCTTCATTCGACGAGGATGTAGAGAAATAGACTTAATTCACATGAG 3355
 |||||||
 729 CCGCGTGAACCTTCATTCGACGAGGATGTAGAGAAATAGACTTAATTCACATGAG 670
 |||||||
 3536 GCGTCATCTCACACCTTAAAGAGGAGATTTAGAGAAAGGAGGACATTTCTTTG 3595
 |||||||
 669 GCGTCATCTCACACCTTAAAGAGGAGATTTAGAGAAAGGAGGACATTTCTTTG 610
 |||||||
 3596 TTCTCCATCATTTTAAATGTGGAGGCTGTTCAGTTTCTTACTTACTATGATAT 3655
 |||||||
 609 TTCTCCATCATTTTAAATGTGGAGGCTGTTCAGTTTCTTACTTACTATGATAT 550
 |||||||
 3656 TCTTCGTACAGTGTCCAAAAAGAAAGCCCAATGAGTGTCTTGAATTTGTTCTT 3715
 |||||||
 549 TCTTCGTACAGTGTCCAAAAAGAAAGCCCAATGAGTGTCTTGAATTTGTTCTT 490
 |||||||
 3716 GATCCCGCATCTTCTTTTATTTAGCATGTGTGCGGCTCTTAATTTGGGTATG 3775
 |||||||
 489 GATCCCGCATCTTCTTTTATTTAGCATGTGTGCGGCTCTTAATTTGGGTATG 430
 |||||||
 3776 AGCAAAATTAACCATGTGTGTGTGCGGCTACCCAGGAGGACTCCCACTTCTGACTGAA 3835
 |||||||
 429 AGCAAAATTAACCATGTGTGTGTGCGGCTACCCAGGAGGACTCCCACTTCTGACTGAA 370
 |||||||
 3836 GTGAGCTGAGAAAGATTCACAGAGTGTATCTGGCCAGATTAAAGATTAATTTCT 3895
 |||||||
 369 GTGAGCTGAGAAAGATTCACAGAGTGTATCTGGCCAGATTAAAGATTAATTTCT 310
 |||||||
 3896 TTGTTCCCTCCCTCCGAGAGACCTCTTATTTATGTGCGGCTCTTATGATTAATTC 3955
 |||||||
 309 TTGTTCCCTCCCTCCGAGAGACCTCTTATTTATGTGCGGCTCTTATGATTAATTC 250
 |||||||

QY 3956 CTTTGATTTGACCTTTGTTGAGAAAGAGGTTGACAGTAGATTAAGCAAGTTCCAGTGCA 4015
 |||||||
 DB 249 CTTTGATTTGACCTTTGTTGAGAAAGAGGTTGACAGTAGATTAAGCAAGTTCCAGTGCA 190
 |||||||
 QY 4016 AATTTACAGTGTGTAGAGTGTGGGGAATTAAGTCTTATTTTCCCTACATGGATA 4075
 |||||||
 DB 189 AATTTACAGTGTGTAGAGTGTGGGGAATTAAGTCTTATTTTCCCTACATGGATA 130
 |||||||
 QY 4076 CAACACTGTGAATTAATTTCAACTGAGAGCCCTGCAGTCTCCCTAAACATAGTTGT 4135
 |||||||
 DB 129 CAACACTGTGAATTAATTTCAACTGAGAGCCCTGCAGTCTCCCTAAACATAGTTGT 70
 |||||||
 QY 4136 TGTGTTTCTTTAACAAGTTTAAAGCTGATTAATAATTAATAAATTTG 4186
 |||||||
 DB 69 TGTGTTTCTTTAACAAGTTTAAAGCTGATTAATAATTAATAAATTTG 19
 |||||||

RESULT 12
 BI820795/c 697 bp mRNA linear EST 04-OCT-2001
 LOCUS 603034002E1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5175097 5',
 DEFINITION mRNA sequence.
 ACCESSION BI820795 GI:15932345
 VERSION BI820795
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 697)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M1436 row: c column: 02
 High quality sequence stop: 693.

FEATURES

source
 1. 697
 /location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5175097"
 /clone_11b="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 6 male brains, age range 23-27; 1
 male lung, age 27; and 1 male testis, age 69. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.8 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research genetics tracking code
 021. Note: this is a NIH-MGC Library."
 BASE COUNT 210 a 167 c 118 g 202 t
 ORIGIN

Query Match 15.7%; Score 685; DB 13; Length 697;
 Best Local Similarity 99.9%; Pred. No. 1.9e-82;
 Matches 696; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2103 TTGAGATTATGAGGCTGATTTGATTTCTTCCCTTTTGGGAATAGGCTGTCAA 2162
 |||||||
 DB 697 TTGAGATTATGAGGCTGATTTGATTTCTTCCCTTTTGGGAATAGGCTGTCAA 638
 |||||||

FEATURES	LOCATION/Qualifiers	EST 21-FEB-2001
2163	GCCTAACCTAATAGGATGGCAGATTCACAGAGTTTACAGGGCTGTCTTATACATTTCCCT	2222
2163	GCCTAACCTAATAGGATGGCAGATTCACAGAGTTTACAGGGCTGTCTTATACATTTCCCT	2222
637	GCTTAAGCTAATAGGATGGCAGATTCACAGAGTTTACAGGGCTGTCTTATACATTTCCCT	578
2223	ATGTTAAAGGGGGTAAAAAGGGCTCTCTTCATTAGACATAGTGGAAATGAAGCAGCCCTTC	2282
577	ATGTTAAAGGGGGTAAAAAGGGCTCTCTTCATTAGACATAGTGGAAATGAAGCAGCCCTTC	518
2283	CTTTAGAGCTGTGCTCTGCATBGCACCTCTTCTCACCCCTGGTACACCCCTTATATGTGGGT	2342
517	CTTTAGAGCTGTGCTCTGCATBGCACCTCTTCTCACCCCTGGTACACCCCTTATATGTGGGT	458
2343	ATAGTGAATTTTAAACCCCTA-AAATTAACAAACAAACCTCACCATGAGCTTTAGAACCGA	2401
457	ATAGTGAATTTTAAACCCCTA-AAATTAACAAACAAACCTCACCATGAGCTTTAGAACCGA	398
2402	AGAGGAATGACAAAGTGAAGCGATGACGAACCAAGCCATCTTACACAGTAGAGAAAAAGACATCG	2461
397	AGAGGAATGACAAAGTGAAGCGATGACGAACCAAGCCATCTTACACAGTAGAGAAAAAGACATCG	338
2462	AGAGTGTGTGATACAGCTGTGAAAGATAGTGTTCATTTGAAACATTCGTGATACCA	2521
337	AGAGTGTGTGATACAGCTGTGAAAGATAGTGTTCATTTGAAACATTCGTGATACCA	278
2522	GTCATGTGGGAAGGATGTTTGGCTGTGATTTATTTTTCAGTTAATGGATTAACATTTCT	2581
277	GTCATGTGGGAAGGATGTTTGGCTGTGATTTATTTTTCAGTTAATGGATTAACATTTCT	218
2582	TTACTGCTCAAAAACCAAAATCTTTGGAAAAAGAAAGTGGGATGTGATTTTCAGAACAA	2641
217	TTACTGCTCAAAAACCAAAATCTTTGGAAAAAGAAAGTGGGATGTGATTTTCAGAACAA	158
2642	GTTTCACAGCTGTAAACCAAAAGCAGCTTATTTTGGGATGGCATGCGCAAAACCTGTATTAAT	2701
157	GTTTCACAGCTGTAAACCAAAAGCAGCTTATTTTGGGATGGCATGCGCAAAACCTGTATTAAT	98
2702	GTCCTTGTATCACATCATCTTCTCAAGTATTTCTTCATTTGGGCTTCATCTTTTAGCAGAA	2761
97	GTCCTTGTATCACATCATCTTCTCAAGTATTTCTTCATTTGGGCTTCATCTTTTAGCAGAA	38
2762	CTCTTGGTGGGATGAGACTTAAGGAGAGGAGGTAGGG	2798
37	CTCTTGGTGGGATGAGACTTAAGGAGAGGAGGTAGGG	1
RESULT 13		
BG289135	864 bp	linear
LOCUS	6023875252P1 NIH_MGC_93 Homo sapiens	EST 21-FEB-2001
DEFINITION	mRNA sequence.	
ACCESSION	BG289135	
VERSION	BG289135.1	GI:13044673
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
TITLE	1 (bases 1 to 864)	
JOURNAL	NIH-MGC http://mhc.nci.nih.gov/	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgaabs-remail.nih.gov	
	Tissue Procurement: ATCC	
	cDNA Library Preparation: Life Technologies, Inc.	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	
	DNA Sequencing by: Incyte Genomics, Inc.	
	Clone distribution: MGC clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LNLN at:	
	http://image.lnl.gov	
	plate: LHAM10407	
	High quality sequence stop: 634.	
	location/Qualifiers	

Query Match	Best Local Similarity	Matches	Conservative	Score	Pred. No.	DB	Length	Indels	Gaps
1972	92.4%	759	0	679.8	8.4e-82	DB 12	864	5	4
<p>Query Match 15.5%; Score 679.8; DB 12; Length 864; Best Local Similarity 92.4%; Pred. No. 8.4e-82; Matches 759; Conservative 0; Mismatches 57; Indels 5; Gaps 4;</p>									
1972	GAGAGTGGCAGACGCTGTATATCTGACGCTTTGCTGCACATATGATATCAGCTGGAT	2031	GAGAGTGGCAGACGCTGTATATCTGACGCTTTGCTGCACATATGATATCAGCTGGAT	60	1	GAGAGTGGCAGACGCTGTATATCTGACGCTTTGCTGCACATATGATATCAGCTGGAT			
2032	AATACCCCTTTCATCTCTGACTTGTGTTTCATTACTCGATTTACAAAACTCTT	2091	AATACCCCTTTCATCTCTGACTTGTGTTTCATTACTCGATTTACAAAACTCTT	61	AATACCCCTTTCATCTCTGACTTGTGTTTCATTACTCGATTTACAAAACTCTT	120			
2092	CATTGGGCTAATTTGTAGTTATGGAGGGTATTTGGGATTTCTTCCCTTTTGGGAAA	2151	CATTGGGCTAATTTGTAGTTATGGAGGGTATTTGGGATTTCTTCCCTTTTGGGAAA	121	CATTGGGCTAATTTGTAGTTATGGAGGGTATTTGGGATTTCTTCCCTTTTGGGAAA	180			
2152	TGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAAAGTTTCAGGGGCTGTTCTA	2211	TGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAAAGTTTCAGGGGCTGTTCTA	181	TGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAAAGTTTCAGGGGCTGTTCTA	240			
2212	TACATTTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGGAAAGTAA	2271	TACATTTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGGAAAGTAA	241	TACATTTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGGAAAGTAA	300			
2272	GCACGCCCCCTTAAAGCTGTGCTCCGTCAGATGGAGACTTCTCACCCTGGTACACCTCC	2331	GCACGCCCCCTTAAAGCTGTGCTCCGTCAGATGGAGACTTCTCACCCTGGTACACCTCC	301	GCACGCCCCCTTAAAGCTGTGCTCCGTCAGATGGAGACTTCTCACCCTGGTACACCTCC	360			
2332	TTAATAGTGGTATAGTATTTTAAACCTTAATAATAAACAACAACCTCACCATGACCTT	2391	TTAATAGTGGTATAGTATTTTAAACCTTAATAATAAACAACAACCTCACCATGACCTT	361	TTAATAGTGGTATAGTATTTTAAACCTTAATAATAAACAACAACCTCACCATGACCTT	420			
2392	TAGAGCAAGAAGAGGATGACAAAGTGAAGGATTAAGCAAGCCATCTTCACAGGTGAA	2451	TAGAGCAAGAAGAGGATGACAAAGTGAAGGATTAAGCAAGCCATCTTCACAGGTGAA	421	TAGAGCAAGAAGAGGATGACAAAGTGAAGGATTAAGCAAGCCATCTTCACAGGTGAA	480			
2452	AAGACATCGAGAGTGGTAGATTAACGTCTGAAAGAAGTATGTTTCAATTTGGAACATTT	2511	AAGACATCGAGAGTGGTAGATTAACGTCTGAAAGAAGTATGTTTCAATTTGGAACATTT	481	AAGACATCGAGAGTGGTAGATTAACGTCTGAAAGAAGTATGTTTCAATTTGGAACATTT	540			
2512	CTGTGATACAGTCAATGTGGGAAGGAGTGTGGCTGTGATTAATTTTCAATTAATGAT	2571	CTGTGATACAGTCAATGTGGGAAGGAGTGTGGCTGTGATTAATTTTCAATTAATGAT	541	CTGTGATACAGTCAATGTGGGAAGGAGTGTGGCTGTGATTAATTTTCAATTAATGAT	599			
2572	AACATTTCTTTACGCTC-AAACAAACCAATCTTTGGAAAAAGAAAGGGAGGTT--	2628	AACATTTCTTTACGCTC-AAACAAACCAATCTTTGGAAAAAGAAAGGGAGGTT--	599	AACATTTCTTTACGCTC-AAACAAACCAATCTTTGGAAAAAGAAAGGGAGGTT--	659			
600	TACATATCTTTACTGCTCAAAAAACCAAAATCTTTGGAAAAAGAAAGGGAGTGTGCA	659	TACATATCTTTACTGCTCAAAAAACCAAAATCTTTGGAAAAAGAAAGGGAGTGTGCA	659	TACATATCTTTACTGCTCAAAAAACCAAAATCTTTGGAAAAAGAAAGGGAGTGTGCA	718			
2629	AGTTTCAGAACAGTACAGCTGTAAACAAAGACACTTATTTGGAGTGGATGCA	2688	AGTTTCAGAACAGTACAGCTGTAAACAAAGACACTTATTTGGAGTGGATGCA	718	AGTTTCAGAACAGTACAGCTGTAAACAAAGACACTTATTTGGAGTGGATGCA	778			
2689	AACCTGTATAATGTCTGTATACATCACTTCTCAAGTATTCCTTCATTTGGGCTTCAT	2748	AACCTGTATAATGTCTGTATACATCACTTCTCAAGTATTCCTTCATTTGGGCTTCAT	778	AACCTGTATAATGTCTGTATACATCACTTCTCAAGTATTCCTTCATTTGGGCTTCAT	838			
2749	CCTTTTAGCAAGACTTGGTGTGGGATAGAGACTTAGGG	2789	CCTTTTAGCAAGACTTGGTGTGGGATAGAGACTTAGGG	838	CCTTTTAGCAAGACTTGGTGTGGGATAGAGACTTAGGG	898			

[illegible]

FEATURES	SOURCE
REFERENCE	BOB76347
AUTHORS	AGENCOURT 8565335
TITLE	IMAGE:6192304 5', mRNA sequence.
JOURNAL	BOB76347
COMMENT	BOB76347.1 GI:22268353
ORGANISM	EST.
SOURCE	human.
LOCUS	Homo sapiens
DEFINITION	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 933)
ACCESSION	NIH-MGC http://mgi.nci.nih.gov/.
VERSION	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
KEYWORDS	Contact: Robert Strausberg, Ph.D.
ORIGIN	Email: cgapbs-remail.nih.gov
LOCUS	Tissue Procurement: Dr. James R. Lupski
DEFINITION	cDNA Library Preparation: Life Technologies, Inc.
ACCESSION	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
VERSION	DNA sequencing by: Agencourt Bioscience Corporation
KEYWORDS	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
ORIGIN	Plate: LHAM13594 row: b column: 17
LOCUS	High quality sequence stop: 508.
DEFINITION	location/Qualifiers
ACCESSION	1. .933
VERSION	/organism="Homo sapiens"
KEYWORDS	/db_xref="taxon:9606"
ORIGIN	/clone="IMAGE:6192304"
LOCUS	/clone_id="Lupski_sympathetic.trunk"
DEFINITION	/sex="male"
ACCESSION	/tissue_type="sympathetic trunk"
VERSION	/dev_stage="adult, 16 yr"
KEYWORDS	/lab_host="DH10B"
ORIGIN	/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dT priming. directionally cloned using the following adaptors: 5'-TCGACCCACGCGTCG-3' and 5'-GACTAGTCTAGATCGAGCGAGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."
LOCUS	185 a 261 c 190 g 296 t 1 others

ORIGIN

Query Match 15.2%; Score 663.2; DB 14; Length 933;
Best Local Similarity 94.7%; Pred. No. 1.3e-79;
Matches 731; Conservative 0; Mismatches 33; Indels 8; Gaps 4;

```
QY 3175 GCCAAGGTATGTAGGCTGTGTGCTGGCCCATCATGTAGAGCAGCAGAGAGATGGGATA 3234
    |||
Db 1 GCCAAGGTATGTAGGCTGTGTGCTGGCCCATCATGTAGAGCAGCAGAGAGATGGGATA 60
    |||
QY 3235 CCATTGTGGGAAGAGAAAAGTTCTCCAGGGGCTCCCACTGCTAAAGTTTTTGTGA 3294
    |||
Db 61 CCATTGTGGGAAGAGAAAAGTTCTCCAGGGGCTCCCACTGCTAAAGTTTTTGTGA 120
    |||
QY 3295 GATGTGATCTGCTCTCGGATTTGACTTTTAAGGAATTAATCTGGCAGCAATGTA 3354
    |||
Db 121 GATGTGATCTGCTCTCGGATTTGACTTTTAAGGAATTAATCTGGCAGCAATGTA 180
    |||
QY 3355 GTATTCTGATGATCTTGTGCTCTTATTTCTCCTTTTGTGTGTGTGTGTGTG 3414
    |||
Db 181 GTATTCTGATGATCTTGTGCTCTTATTTCTCCTTTTGTGTGTGTGTGTGTG 236
    |||
QY 3415 TGGCTATGGGTTTCACTTTGTAATCTCATCTGCTTAGAGAGTGGGCTCTATAGGA 3474
    |||
Db 237 TGGCTATGGGTTTCACTTTGTAATCTCATCTGCTTAGAGAGTGGGCTCTATAGGA 296
    |||
QY 3475 ACCGTGTAACCTTCATTCAGCAGAGATGTAGAGAAATAGGACTTAATCCACTAG 3534
    |||
Db 297 ACCGTGTAACCTTCATTCAGCAGAGATGTAGAGAAATAGGACTTAATCCACTAG 356
    |||
QY 3535 GGGCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACCTGGCCAGATTTCTTT 3594
    |||
Db 357 GGGCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACCTGGCCAGATTTCTTT 416
    |||
QY 3595 GTTCTCATCTTTTAAATGTGGCAGGCTTTCACTTTCTTACTTACTATGTGATAT 3654
    |||
Db 417 GCTCTCATCTTTTAAATGTGGCAGGCTTTCACTTTCTTACTATGTGATAT 476
    |||
QY 3655 TTCTTGTAACTGTGCAAAAAGAAAAGAACCCATCAGTGTCTGACTTGTCTT 3714
    |||
Db 477 TTCTTGTAACTGTGCAAAAAGAAAAGAACCCATCAGTGTCTGACTTGTCTT 536
    |||
QY 3715 TGATCCGTCAGTTCTTCTGATTTTCAGCATGTGCGGGTCTAAATTTGGTATGAGT 3774
    |||
Db 537 TGATCCGTCAGTTCTTCTGATTTTCAGCATGTGCGGGTCTAAATTTGGTATGAGT 596
    |||
QY 3775 TAGCAATTTAACCATGTGTGTGTGCGCTAACCCAGAGGG--ACTGCCAGTTTCTGACTT 3832
    |||
Db 597 TAGCAATTTAACCATGTGTGTGTGCGCTAACCCAGAGGGACTCCCGAGTTTCTGACTT 656
    |||
QY 3833 GAATGAGCTGAG--AAGAAATCCAGAGTGTATCTGCGCAGATTTAAGTATCTATT 3891
    |||
Db 657 CACGACACCGAGAAATACCCCTACAGTGTCTCCCGCCGCACTAGATGATTTATT 716
    |||
QY 3892 TCCCTGGTCTCCCTCCCTGAGAGACTC-TTATTTATTTGTCCTCTTC 3942
    |||
Db 717 TCCCTGGGCTCCCTCTCCCTGAGAGACTCTTTACTTATTTGTCCTCTTC 768
    |||
```

Search completed: July 14, 2003, 13:38:26
Job time : 5657 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 19:15:06 ; Search time 370 Seconds

(without alignments)
3737.100 Million cell updates/sec

Title: US-10-054-935-2
3338
Perfect score: 1 MTMRSAVFKAAPAGNPPE.....RSRCRLFIQKQPHRTCRK 614
Sequence:

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlh
-MODEL-frame+ p2n model -DB=us10054935/runat_14072003_115025_24619/app.query.fasta.1.775
-Q=/cgrn21/USPRO.spool/US10054935/runat_14072003_115025_24619/app.query.fasta.1.775
-DB=us10054935 -Q=us10054935 -START=1 -END=1 -MATRIX=blomus62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000
-USER=us10054935.ecgn.1.1.263 &runat_14072003_115025_24619 -NCPU=6 -ICPU=3
-NO_MMAR -LARGOQUERY -NEG_SCORES=0 -WAIT -DSRBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :
1: N Geneseq/101002: *
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1448.5	44.7	1529	21 AAF18193	Lung cancer associ
2	1351.5	41.7	2887	21 AAF6874	Human ORFX ORF2429
3	1055.5	32.6	615	24 ABK62818	Rat sequence diffe
4	680.5	21.0	2887	21 AAC76874	Human ORFX ORF2429
5	246	7.6	7713	22 AAK51958	Human polynucleoti
6	245	7.6	5373	24 AAD30567	Human kinase polyp
7	238	7.4	9551	20 AAZ22301	CDNA encoding a hu
8	236	7.3	7033	21 AAZ36989	Human polynucleoti
9	235.5	7.3	4037	22 AAK52942	Human polynucleoti
10	231	7.1	3608	24 AAD33243	Human secreted pro
11	228.5	7.1	3169	22 AAI58822	Human polynucleoti
12	228.5	7.1	3203	22 AAI60608	Human polynucleoti
13	225	6.9	2108	24 ABL67774	Oesophagus cancer
14	223.5	6.9	3117	21 AAZ38858	Human Jurkat cell
15	223.5	6.9	3414	23 AAS78091	Human encoding novel
16	222.5	6.9	8459	22 AAC89557	Human histone deac
17	222.5	6.9	117213	19 AAV62176	HSV-2 strain SB5 C
18	221.5	6.8	1833	23 AAS67931	DNA encoding novel
19	221.5	6.8	7516	22 AAK51987	Human polynucleoti
20	220.5	6.8	8460	24 ABK87718	Human cDNA encodin
21	220	6.8	2438	24 ABO54916	Human ovarian anti
22	220	6.8	6432	23 AAS65903	DNA encoding novel
23	220	6.8	13555	22 AAS34526	Human DNA for a no
24	220	6.8	13555	22 AAL04593	Human reproductive
25	220	6.8	13555	23 ABL97516	Human testicular a
26	220	6.8	154746	24 AAD25519	Human herpesvirus
27	220	6.7	154746	24 AAD25519	Human herpesvirus
28	218	6.7	6457	24 ABN96892	Gene #3390 used to
29	217.5	6.7	2614	23 ABL28430	Drosophila melanog
30	217.5	6.7	4954	23 ABL05187	Drosophila melanog
31	217.5	6.7	9516	23 ABL05187	Drosophila melanog
32	217	6.7	3464	23 ABL30199	Drosophila melanog
33	217	6.7	3959	24 AB199206	Mouse ischaemic co
34	217	6.7	5524	23 ABL30198	Drosophila melanog
35	217	6.7	36778	21 AAZ87318	S. venezuelae pik
36	217	6.7	37948	21 AAZ87318	S. venezuelae pik
37	217	6.7	38506	21 AAZ75633	Nucleotide sequenc
38	217	6.7	38506	21 AAZ56001	Recombinant cosmid
39	216.5	6.7	3720	23 AAS74064	DNA encoding novel
40	216	6.7	33023	24 ABN96853	Gene #3351 used to
41	215.5	6.7	3147	22 AAS03399	Thermus DNA encodi
42	215	6.6	14835	24 AAS94858	Human DNA sequence
43	214.5	6.6	7746	21 AAA07836	Human homologue of
44	214.5	6.6	4411529	22 AAI99682	Mycobacterium tube
45	213.5	6.6	2725	24 ABK84458	Human cDNA differe

ALIGNMENTS

RESULT 1
ID AAF18193 standard: DNA; 1529 BP.
AAAF18193

14-MAR-2001 (first entry)

Lung cancer associated polynucleotide sequence SEQ ID 212.

Human; lung cancer associated protein; neuroprotective; cytostatic;

cardioactive; immunomodulatory; muscular active; vulnerable;

gastrointestinal; nephrotoxic; anti-infective; gynecological;

antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

proliferative disorder; wound healing; infectious disease; ds.

Homo sapiens.

PN W0200055180-A2.
 XX 21-SEP-2000.
 PD 08-MAR-2000; 2000WO-US05918.
 XX 12-MAR-1999; 99US-0124270.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C. A.
 XX Ruben SM;
 XX WPI: 2000-587514/55.
 DR P-PSDB: AAB58317.
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 676-677; 1425pp; English.
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytosolic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SO Sequence 1529 BP; 460 A; 344 C; 379 G; 344 T; 2 other;
 Alignment Scores:
 Pred. No.: 6,45e-43 Length: 1529
 Score: 1448.50 Matches: 284
 Percent Similarity: 95.64% Conservative: 1
 Best Local Similarity: 95.30% Mismatches: 6
 Query Match: 44.73% Indels: 7
 DB: 21 Gaps: 2
 US-10-054-935-2 (1-614) x AAF18193 (1-1529)
 QY 181 ProProLeuAlaProThrAlaThrAlaGlyThreLeuAlaAlaSerGluGlyArgTrpLys 200
 DB 3 CCACCCCTCGCGCCACCGCCACCGCGGACCTGCGCGCGACGAGGAGATGGAAG 62
 QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 220
 DB 63 AGTATGAGAAAGACCTCTCTGCGGGGTGTCGCGGCTCGGAGCCTCCAGTCAGCGCGCC 122
 QY 221 CysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGluGlnGlnGlnGlnLeu 240
 DB 123 TGCCTCAAAACAGATCCTCTGCTGCTCAATTGCACTTCATCGAACGACGACACACCTG 182
 QY 241 GlnAlaLysGluLysGluIleGluGluLeuLysSerGluArgAspThrLeuAlaArg 260
 DB 183 CAGGCCAAGGAAAGGAGATCGAGAGCTGAGTCAAGAGAGACAGCTCTCTGCTCGG 242
 QY 261 IleGluArgMetGluArgMetGlnLeuValLysLysAspAsnGluLysGluArgHis 280
 DB 243 ATTGAAGTATGGAAGCGGATGCTGCTGTAAGAAGATTAACGAAAGAAAGGCAC 302

QY 281 LysLeuPheGlnGlyTrpGluThrGluLysArgGluGluThrGluLeuSerGluLysIle 300
 DB 303 AAGCTGTTTCAGGGCTATGAAACTGAGAGAGAGAGAAACAGAGTATCTGAGAAATT 362
 QY 301 LysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProPheProPhe 320
 DB 363 AAACGTGAGTGCACCGCGACCTTCGAGACATCCAGACTGCTGCCCAACCCCTTC 422
 QY 321 SerCysGluArgSerGlyLysGlyHisLysArgLysSerProPheLysSerThrGluArg 340
 DB 423 TCATGTGGCGCGAGTGGAAAGGACATGAAAGAAATCCCATTTGGAATGAGAAAGA 482
 QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys 360
 DB 483 AAGACTCCTGTTAAAGAGCTGCTCGAATTTCAAAAGCAAAACAAACCTCTAAG 542
 QY 361 HisSerProIleLysGluGluProCysGlySerLeuSerGluThrValLysArgGlu 380
 DB 543 CACTCTCTTAAAGAGGAAACCTGTGTCTTATCTGAAACTGTTTGAACGTGAA 602
 QY 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProArgLeu 400
 DB 603 TTGAGAGCCAGAAAGAACCCAGAAAGCCCGCTTCAGTGAGACCCACAGAGCTC 662
 QY 401 SerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIle 420
 DB 663 TCCACTCCCAAAAGGAGACCCAGACCCATCCCAAGAGAAAGCCTCTCAAGAGATGA 722
 QY 421 GluAspLeuProThrLysSerThrThrGluMetThrLeuLysArgTrpHisGlnProPro 440
 DB 723 GAACATTTGCCGATCTTTCCACCCAGAAATGATTTGTCTTGCCACCCAGCTGCC 782
 QY 441 ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys 460
 DB 783 CCATCACCAGTATACCATTCAGGGAATCTCTCCAAAGAGAGAGAGATGATGAC 836
 QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
 DB 837 -----ACTAAGCATAGAGAAACACTTGTCTT---ATACCTTAGTGG 875
 RESULT 2
 ID AAC76874
 XX AAC76874 standard; cDNA; 2887 BP.
 AC AAC76874;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORF2429 polynucleotide sequence SEQ ID NO:4857.
 XX
 KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
 KW vulnary; antiproliferative; antiparkinsonian; neurotoxic; neuroprotective;
 KW anticonvulsant; osteoprotic; antitachycardic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42665.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4044-4045; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antipneumatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteoplastic; anticoagulant; antithrombotic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antitumor; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy.
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypohyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other:

Alignment Scores:
 Pred. No.: 2,62e-39 Length: 2887
 Score: 1351.50 Matches: 269
 Percent Similarity: 94.74% Conservative: 1
 Best Local Similarity: 94.39% Mismatches: 8
 Query Match: 41.74% Indels: 8
 DB: 21 Gaps: 2

US-10-054-935-2 (1-614) x AAC76874 (1-2887)

QY 194 AlAserGIuGIaGrTrpLysSerMetArgLysSerProLeuGIuGIyGIyGIySer 213
 DB 5 GCCGGCAGGGGACAGATGGAAGATGATGAGAGAGCCCTCCGGGGTGGGGGGCTCG 64
 QY 214 GIyAlAserSerGIuAlaIaCysLeuLysGIuLLeuLeuLeuGIuLLeuValLys 233
 DB 65 GGAGC-TCGAGTCCGGCCGGCCCTCAACAGATCCCTTCGTCGAATTGGACCTCATC 123
 QY 234 GIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 253
 DB 124 GAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183
 QY 254 ArgAspThrLeuLeuAlaIaCysLeuLysGIuLLeuLeuLeuGIuLLeuValLys 273
 DB 184 AGAGACACGCTCTTCGTCGATGGAAGATGGAAGAGCGGAGCAGCTGTAAAGAG 243
 QY 274 AspAsnGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 293
 DB 244 GATACGAGAAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 303
 QY 294 ThrGIuLeuSerGIuLysIleLysLeuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 313

DB 304 ACAAGCTATCTGAGAAATTAACCTGAGTGCAGCCGAGCTTCGAGACATCCAG 363
 QY 314 ThrLeuProProLysProPheSerCysGIyArgSerGIyGIyGIyGIyGIyGIySer 333
 DB 364 ACTCTGCTCCCAAGCCCTTCATGTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
 QY 334 ProPheGIySerThrGIuArgLysThrProValLysLysLeuAlaProGIuPheSerLys 353
 DB 424 CCATTGGAGCTACAGAAAGAGACCTCTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
 QY 354 ValLysThrLysThrProLysHisSerProIleLysGIuGIuGIuGIuGIuGIuGIuGIu 373
 DB 484 GTCAAAACAAAACTCTTAAGCACTCTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 543
 QY 374 GIuThrValLysLysArgGIuLeuArgSerGIuGIuGIuGIuGIuGIuGIuGIuGIu 393
 DB 544 GAAACGTGTGTAACGTGAATTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 603
 QY 394 ValAspThrProProArgLysSerThrProGIuLysGIuGIuGIuGIuGIuGIuGIuGIu 413
 DB 604 GTGAGACACCCACCAAGACTCTCCACTCCCAAAAGGAGGAGGAGGAGGAGGAGGAGGAG 663
 QY 414 LysAlaPheSerSerGIuLLeuGIuAspLeuProTyrLeuSerThrThrGIuMetTyrLeu 433
 DB 664 AAACCTTCATCACTGATGATGAGATTTGGCTTCCACACAGAAATGATTTG 723
 QY 434 CysArgTrpHisGIuProProProSerProLeuProLeuArgGIuSerSerProLysLys 453
 DB 724 TGCTGTGGGACCCAGCTCCCTCCACCGTACCATTTAGCGGGAATCTCTCCAAAGAG 783
 QY 454 GIuGIuThrValAlaIaCysLeuMetProSerSerValAlaGIuGIuThrSerValLeu 473
 DB 784 GAGGAGACTGTGAGA-----AGTAAAGCATAGAGAACCTTGCTTT 825
 QY 474 AlaValProSerTrp 478
 DB 826 ---ATACCTAGTGG 837

RESULT 3

ABK62818 ID ABK62818 standard; cDNA; 615 BP.

XX AC ABK62818;

XX AC 18-JUN-2002 (first entry)

XX DE Rat sequence differentially expressed in response to a hepatotoxin #725.

XX XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 XX XX differential expression; centrilobular necrosis; steatosis.

XX OS Rattus norvegicus.

XX PN W0200210453-A2.

XX PD 07-FEB-2002.

XX PF 30-JUL-2001; 2001W0-US23872.

XX PR 31-JUL-2000; 2000US-222040P.

XX PR 02-NOV-2000; 2000US-244880P.

XX PR 11-MAY-2001; 2001US-290029P.

XX PR 15-MAY-2001; 2001US-290645P.

XX PR 22-MAY-2001; 2001US-292336P.

XX PR 06-JUN-2001; 2001US-295798P.

XX PR 13-JUN-2001; 2001US-297457P.

XX PR 19-JUN-2001; 2001US-298884P.

XX PR 09-JUL-2001; 2001US-303455P.

XX PA (GENE-) GENE LOGIC INC.

XX PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

XX WPI: 2002-241625/29.

XX Predicting toxic effects of compounds or the progression of these toxic
PT effects by determining the changes in gene expression in tissues or
PT cells exposed to the toxin and comparing these to gene expression in
PT unexposed tissues or cells.

PS Claim 1: Seq ID No 725; 239pp; English.

XX The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic
CC effect of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression.
CC The method can also be used to identify an agent which modulates the
CC toxic response and predict cellular pathways that a compound modulates
CC in a cell. The methods utilize a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridises to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used in the specification,
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity
CC is characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent.

XX Sequence 615 BP; 84 A; 235 C; 224 G; 71 T; 1 other;

Alignment Scores:
Pred. No.: 1,63e-29 Length: 615

Score: 1055.50 Matches: 197

Percent Similarity: 97.55% Conservative: 2

Best Local Similarity: 96.57% Mismatches: 4

Query Match: 32.60% Indels: 1

DB: 24 Gaps: 1

US-10-054-935-2 (1-614) x ABR62818 (1-615)

QY 13 A1AProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
DB 3 GCCCTCGCGGCGGCAACCCGAGCAGCAGCAGCAGCGGCTGCGGCTCGGAGC 62
QY 33 GlyProGluAspGluProGlyAlaAlaGlnAlaHisPheLeuProArgHisArgGlyLeu 52
DB 63 GGGCCGAGAGAGAGACTCGGGGCGGCGGAGCCCTCTCTCCCGGCACTGTAAGCTC 122
QY 53 LysGluProGlyProProLeuAlaSerSerGlnGlyGlySerProAlaProSerProAla 72
DB 123 AAGGAGCCCGGCGGCGGCTGCTCTACCCAGGCGGAGCGCGGCGGCTCTCCACT 182
QY 73 GlyCys---GlyGlyLysGlyArgGlyLeuLeuProAlaGlyAlaAlaProGlyGln 91
DB 183 GGGCTCGCGGCGGCGGAGGCGGCGGCTTGTACTCCGCGGCGGCGGCGGCGGCGAG 242
QY 92 GlnGlnGlnSerTyrGlyGlySerValProLeuProCysProProProAlaThrLysGln 111
DB 243 CAGGAAGAGAGACTGGGCGGCTTGCGCTTGGCCCTGTCTCCGCCCGGCTACCAACAA 302

QY 112 AlaGlyIleGlyGlyGluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyr 131
DB 303 GCCCGCATCGGCGGCGGAGCCAGTCGAGCGGCGGCTGCTGCAAGCCCGGCAACTAT 362
QY 132 GlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGluProThrPro 151
DB 363 CAGCGGTGCTGCCCATTCAGACGGGCTCTCTGCGGCGGCGGCAAGAGCTACGAGCC 422
QY 152 TrpAlaGlyAspLysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGly 171
DB 423 TGGCGTGGGAGACAAGGTGGGCGGCTCCCGCAGCTGCCAGCCCTGGAGCCGCGGGA 482
QY 172 ProProProLeuProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThr 191
DB 483 CCCCACACACTACTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 542
QY 192 LeuAlaAlaSerGluGlyArgTyrPheLysSerMetArgLysSerProLeuGlyGly 211
DB 543 CTGGCGGCGGAGTAGGCGGAGATGAGAGATATAGAGAGAGCCCTCTCGGGGTGCGGCG 602
QY 212 GlySerGlyAla 215
DB 603 MGCTCGGAGCC 614

RESULT 4

ID AAC76874/c
AAC76874 standard; cDNA; 2887 BP.

AC AAC76874;

DT 08-FEB-2001 (first entry)

DE Human OREF429 polynucleotide sequence SEQ ID NO:4857.

XX Human; open reading frame; OREF; detection; cytosolic; hepatotropic;
KW vulnerrary; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
KW anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineumatic; antihistaminic;
KW antineumatic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive; ss.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI: 2000-602362/57.

DR P-Psdb; AABA2665.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -

PS Claim 5: Page 4044-4045; 5507pp; English.
XX AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB4397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
CC osteoprotic; anticonvulsant; antiallergic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertensive; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy.
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;

Alignment Scores:
Pred. No.: 7,04e-16 Length: 2887
Score: 680.50 Matches: 134
Percent Similarity: 87.74% Conservative: 2
Best Local Similarity: 86.45% Mismatches: 12
Query Match: 21.02% Indels: 7
DB: 21 Gaps: 2

US-10-054-935-2 (1-614) x AAC76874 (1-2887)

OY 324 ArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrPro 343
DB 2491 AGAGGGGCTGGAGGGGCTTAGACCTTACCCCATTTGGAGAGCAAGAAAGAAAGCTCT 2432
OY 344 ValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSerPro 363
DB 2431 GTTAAAGAGCTGGCTCTGAAATTTTCAAAAGCAAAACAAACCTCTAGACACTCTCT 2372
OY 364 IleLysGluGluProGlySerLysSerLysThrLysLysArgGluLeuArgSer 383
DB 2371 ATTAAAGAGAACCTGTGGTCTTATCTGAACCTGTTTGAACGTAATGAGGAGC 2312
OY 384 GluGluThrProGluLysProArgSerSerValAspThrProProArgLysSerThrPro 403
DB 2311 CAAGAAACCCCAAGAAACCCCGGCTTCAGTGGACACCCCAAGACCTCCACTGCC 2252
OY 404 GluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeu 423
DB 2251 CAAAGAGGACCCAGACCAATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAAGATTGG 2192
OY 424 ProGlyLysSerThrThrGluMetGlyLeuGlySerArgGlyThrPheGlnProProSerPro 443
DB 2191 CGTACCTTTCACACAGAAATGATTTGTGTCTTGGACACCACTCCCAATCACCG 2132
OY 444 LeuProLeuArgGluSerSerProLysGluGluThrValAlaArgCysLeuMetPro 463
DB 2131 TTACCATTAAGGGAATCTCTCCAAAGAGGAGGAGACTGTAGCA----- 2087
OY 464 SerSerValAlaGlyGluThrSerValLeuAlaValProSerThr 478
DB 2086 ---AGTAGGCATAGAGAACCTTGTCTT---ATACCCCTAGTGG 2048

RESULT 5
AAK51958
ID AAK51958 standard; cdna: 7713 BP.
XX
AC AAK51958;

XX 06-NOV-2001 (first entry)
DT
XX Human polynucleotide SEQ ID NO 503.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QJ, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPT: 2001-476283/51.
DR P-PSDB: AAM78825.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 1809-1815; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

Sequence 7713 BP; 2077 A; 1974 C; 2094 G; 1568 T; 0 other;

Alignment Scores:
Pred. No.: 2.3 Length: 7713
Score: 246.00 Matches: 182
Percent Similarity: 32.73% Conservative: 88
Best Local Similarity: 22.06% Mismatches: 289
Query Match: 7.60% Indels: 266
DB: 22 Gaps: 33

US-10-054-935-2 (1-614) x AAK51958 (1-7713)

OY 11 AlaAlaAlaProAlaGlyLysAsnProGluGlnArgLeuAspArgGluArgAlaAla 30
DB 4345 GCTGCGGACGACGAGGCGG-----CCGCTCCGCTAGGCGAGCTGTACGA 4392
OY 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuProArgHis 49

Oy		549	cgluValThrSerPheProGluProAspSp-----ValGIusSerLeuMetIleTh	567
Dd		6514	TGCCACACCACCAAAAGCAGAGCGAAGAAGAACGACGAACAAGAGAGCCCAAGMACGA	6573
Oy		567	rPrOheUeuProValAlaAlaheglYAcrProLeuProLysLeuThrProGlnAsn	587
Dd		6574	GGCCTCAAAGCCCAAGGTGGAGGAGAGAAAGAAAGAAACCTGCTCTCGAAAAAGCCCAAGANT-	6632
Oy		587	heglUeuProTrPLeuAspGlUArSerArGcYsaTgLeuGluIleGlnLysLysGlnT	607
Dd		6633	-----CCAAGCTTGAAAGCCACAGA--AGGAAGAGCGCTGAATATAAGAAAAGATCCC	6681
Oy		607	hrProHisArg	610
Dd		6682	CACCCACAGAGA	6692
RESULT 6				
AAD30567				
ID	AAD30567	standard; cDNA;	5373 BP.	
XX	AAD30567;			
XX	21-MAY-2002	(first entry)		
DE	Human kinase polypeptide (PKIN-20)	cDNA.		
KW	Human; kinase polypeptide; PKIN-20; gene therapy; Addison's disease;			
KM	Leukemia; immune disorder; lymphoma; melanoma; developmental disorder;			
KW	acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;			
KW	asthma; Crohn's disease; rheumatoid arthritis; burstits; atherosclerosis;			
KW	cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;			
KW	cholestasis; cardiacnt; cardiovascular disorder; Niemann-Pick's disease;			
KW	lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;			
KM	drug screening; transgenic animal; antinflammatory; hepatotropic;			
KX	hypotensive; anti-HIV; enzyme; ss.			
OS	Homo sapiens.			
XX				
FH	key	Location/Qualifiers		
FT	CDS	1..4719		
FT		/tag= a		
FT		/product= "Human PKIN-20"		
FT	sig_peptide	1..111		
FT		/tag= b		
FT	mat_peptide	112..4716		
FT		/tag= c		
PN		/product= "Mature human PKIN-20"		
XX	MO200208399-A2.			
PD	31-JAN-2002.			
PF	20-JUL-2001; 2001WO-US23092.			
XX				
PR	21-JUL-2000; 2000US-22003BP.			
PR	28-JUL-2000; 2000US-222112P.			
PR	04-AUG-2000; 2000US-222831P.			
XX	11-AUG-2000; 2000US-224739P.			
PA	(INCY-) INCYTE GENOMICS INC.			
PA	(THOR/) THORNTON M.			
PI	Thornton M, Yue H, Khan FA, Gurrurajan R, Hafalia AJA, Walla NK;			
PI	Patelerson C, Ramkumar J, Gandhi AR, Pollocky JL, Baughn MR;			
PI	TriBouley CM, Bandman O, Nguyen DB, Lu Y, Burford N, Lal P;			
PI	Ding L, Yao MG, Elliott VS, Reipon SA, Kearney LJ, Lu DAM;			
PI	Greenwald SR, Tang YT, Xu Y, Walsh RT, Gietzen KJ, Yang J;			
XX	Hillman JL;			
XX				
DR	WPI; 2002-206083/26.			
DR	P-PSDB; AAE19162.			

New human kinase polypeptide, useful in diagnosis, prevention and treatment of cancer, immune disorder, growth and developmental disorder, cardiovascular disorder and lipid disorder
 Claim 5; Page 194-196; 196pp; English.

The present invention relates to an isolated human kinase polypeptide (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is useful for diagnosing, treating and preventing cancer (e.g., leukemia, lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's disease, rheumatoid arthritis), a growth and developmental disorder (e.g. bursitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial infarction), and a lipid disorder (e.g., fatty liver, cholestasis, gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of drug screening techniques and to analyse the proteome of a tissue or cell type. PKIN is useful for creating knockin humanised animals or transgenic animals to model human diseases, in somatic or germ-line gene therapy, to generate a transcript image of a tissue or cell type, for detecting differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier or affected individuals, and as hybridisation probes for mapping naturally occurring genomic sequences. PKIN is useful in southern or northern analysis, dot blot or other membrane-based technologies, in PCR technologies, in dipstick, pin, multiformat enzyme linked immunosorbent (ELISA)-like assays and in microarrays utilising fluids or tissues from patients to detect altered PKIN expression. The present sequence is human PKIN-20 cDNA.

Sequence 5373 BP; 1037 A; 1636 C; 1812 G; 888 T; 0 other:

Alignment Scores:

Pred. No.:	1.87	Length:	5373
Score:	245.00	Matches:	139
Percent Similarity:	33.07%	Conservative:	68
Best Local Similarity:	22.20%	Mismatches:	199
Query Match:	7.57%	Indels:	222
DB:	24	Gaps:	24

US-10-054-935-2 (1-614) x AAD30567 (1-5373)

```

OY      11 ALaLaLaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluAValAlaAla 30
Db      1248 GCGGAAGCTCCAGTCTGTGA-----GCAGGA 1274
          ||| ||||| |||
OY      31 LeuGlyGlyProGluAspGluPro-----GlyAlaAlaGluAlaHisPheLeuProArg 48
Db      1275 GAAGGTGAGCTGACAGCAAGACCAAGAGGCCCTGCACGCCCCAC----- 1322
          ||||| ||| ||| ||| |||
OY      49 HisArgIleuLeuGluProGlyProProLeuAlaSerSerGlnGlySerProAla 68
Db      1323 -----AGACCATCGGAGAGCTGAGCAGCTACGAGGAAGTGCAGACTCT 1367
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      69 ProSerProAlaGlyCysGlyGlySerGlyLeuArgGlyLeuLeuLeuProAlaGlyAlaAla 88
Db      1368 GCGGAGACAGCTGCCAGAGATGCTGAGGAGCAAGGCCCTCATTTGCCAGACGATGGCC 1427
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      89 ProGlyGlnGlnGluGluSerTyr-----GlyIleSerValProLeuProCysPro 105
Db      1428 CCC-----AGCTGGTAGCCCGACGATGACAGACAGTACGCGAGAGACT 1472
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      106 ProProAlaIleTyrGlnAlaGlyTyrIleGly-----GlyGluProAlaAlaAlaGlyAlaGly 124
Db      1473 TGACCGACTTACGCGGAGCTGGCGAGAGCGGCGAGGCTGCAGAGCTCCAGGACAGGA 1532
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      125 CysSerProArgProTyrGlnAlaValLeuProIleGlnGlySerLeuValAla 144
Db      1533 GCTCTGCAGGGCCCA-----GGGCGACAGGAGGA 1562
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      145 AlaAlaIuGluProThrProTyrAlaGlyAspIleGlyAlaAlaAlaSerProAlaAla 164
Db      1563 GCTGCTTACAGAGCTACAGAGGCCCGACAGAGAGAGAGAGCGCGCACACGACTGACGAGCCG 1622
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```



```

OY 165 ThrAlaSerAspProAlaGlyProProPro-----174
Db 1623 GGCCTTGAGCTCCCTGGAGAGAGCCCGGGCTGCCAGAGAGAGCTGAGGCCAGT 1682
OY 175 ---LeuProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAla 193
Db 1683 GTCCTCCCTGAGCCG-----GCAGGTGACGACGCTGACGAGGACAGTGGGA 1727
OY 194 AlaSerGluGlyAlaTyrTyrPylSerMetArgLysSer-----ProLeu 207
Db 1728 GCAACGCGCTTG-AGGAG-TGCTCCAGGCAAGACCCATCCACACAGCCCTCTGAGACCAAC 1785
OY 208 GlyGlyGlyGlyGlySerGlyAlaSerSerGlyAlaAlaCysLeuGlyGlyLeuLeu 227
Db 1786 GGAATGGGACCCCTGAGGGGTGGCTCCAGAGAGGCCCACTGAGAGAGAGAGTGGCCGCC 1845
OY 228 LeuGlnLeuAspLeuIleGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
Db 1846 CTGGCAGAGCAGCTGAGGAGGCGCCACAGCCACAGCCAGCTGTAAAGAGAGAGGCTCTG 1905
OY 248 GlnGlnLeuGlySerGluAlaTyrThrLeuLeuAlaArgIleGluAlaTyrMetGluAla 267
Db 1906 TGCCAGCTGCGAGAGAGAAACCGAGGCTGAGCGCGGAGCAGAGAGCGCTAGAA----- 1959
OY 268 MetGlnLeuValLysLysAspAsnGlnLysGluAlaTyrGlnLysLeuPheGlnGlyTyrGlu 287
Db 1960 GCAGAGCTGGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2004
OY 288 ThrGlnGluAlaGlnGluThrGlu-----295
Db 2005 -----GAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2058
OY 296 -----2058
Db 2059 TGGGTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2118
OY 302 LeuGlnGlyGlnProGlnLeuSerGlnThrSerGlnThrLeuProPylSerProPheSer 321
Db 2119 GAGAGCTGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2178
OY 322 CysGlyAlaTyrSerGlyLysGlyLysAlaTyr-----LysSerPro 334
Db 2179 ATGAGAGCTCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2238
OY 335 PheGlySerThrGlnGlyLysThrProValLysLys-----346
Db 2239 CAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 346
OY 347 -----346
Db 2299 CTGCAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 356
OY 357 LysThrProLysHisSerProIleLysGlnGluProCysGlySerLeu-----372
Db 2359 GAGCTCGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
OY 373 -----SerGlu-ThrValCysLysArgGlnLeuArgSerGlnGlnGlnGlnGlnGln 2418
Db 2419 TCCTTCGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2478
OY 388 u-----388
Db 2479 AGGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 388
OY 389 ---LysProArgSerSerValAspThrProProArg-----399
Db 2539 GTGTTCCCAAGAGACCCAGCTGCCACACAGCTCTACAGAGAGAGAGAGAGAGAGAGAG 399
OY 400 -----LysSerThrProLysGlyPro-----407
Db 2599 TGGGCGATGAGGCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
OY 408 -----SerThrHisProLysGlnLysAlaPheSerSerGlnIleGln 421

```

```

Db 2659 CACAGCTGCGCCCGGAGAGCTTCCCATCCCCGA-----2692
OY 421 uAspLeuProTyrLeuSerThrThrGlnMetTyrLeuCysArgTyr-----436
Db 2693 -----CCAAGTGTCTCGGCTGACCTCGGTGATGCTGGAGCTGGGCGCCAGGCGCTG 2745
OY 437 -----HisGlnProProProse 442
Db 2746 GGTGTGATGCTCGGCTGACTTTTGTACACAACTGTGCTCCACAGAGCGCCACCTGCG 2805
OY 442 rProLeuProLeuArgLysSer-----SerProLysGlnGlnGlnGlnGlnGln 456
Db 2806 CCGGTCGCCCTGACCTCTCCGACAGAGCCCTGGAGAGTACCCCGAGAGAGAGAGAGAG 2865
OY 456 rVal-----AlaArgCysLeuMetProSerSerValAlaGlyGluThrSe 471
Db 2866 ACTGCTATGAGGCTTTCTGTGCTGCGCGGCGCTCAGGTGTCCGCGGCGGCTGGCAG 2925
OY 471 rValLeuAlaValPro 476
Db 2926 CCGGTGTGTGCTGCC 2941

```

RESULT 7

AAZ22301 standard; cDNA; 9551 BP.

AAZ22301;

25-NOV-1999 (first entry)

cDNA encoding a human trichohyalin (TRHY) protein.

Human: trichohyalin; TRHY; protein; tissue structure; wound healing; terminal differentiation; epidermal tissue; proteinaceous gel;

breast implant; ss.

Homo sapiens.

US9598752-A.

28-SEP-1999.

14-FEB-1997; 97US-0800644.

30-APR-1993; 93US-0056200.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Kim I, Chung S, Park S, Steinert PM, Lee S;

WPI; 1999-561041/47.

P-PSDB; AAY30795.

Human trichohyalin useful for forming a proteinaceous gel that promotes wound healing -

Claim 1; Fig 3A-W; 126pp; English.

The present sequence encodes a human trichohyalin (TRHY) protein.

The protein is found in terminally differentiating epidermal tissue,

and is involved in forming the structural architecture of such

tissue. The trichohyalin protein is useful for forming a

proteinaceous gel which may then be used for healing wounds, or in

breast implants.

Sequence 9551 BP; 2907 A; 1974 C; 2870 G; 1800 T; 0 other;

Alignment Scores:

```

Pred. No.: 5.21
Score: 238.00
Percent Similarity: 35.67%
Best Local Similarity: 23.68%
Length: 9551
Matches: 162
Conservative: 82
Mismatches: 265

```


Dh 1496 CCT-GGACGACAGTCTGACACATGCGGG-AGGTGGCCCGACGCGGACAGCTGGAG 1553
Qy 341 sThpProValylsLeuAlaProGluPheSerValylsThr----- 358
Dh 1554 GTGGAGCATGACGAGGCTCGGCTAGCCTACGGGAGGAGGAGTCCGGAGACTG 1613
Qy 359 -----ProLysHisSerPro-----IleLysGluGluProCysGly-----SerLe 372
Dh 1614 CAGCAGGCCAGGCTGAAGCCCAAGAGAACATAGAGAGCGGCTCTGGAGTCT 1673
Qy 372 uSerGluThrValCysLysArgGluLeu-ArgSerGluGluThrProGluLysProArg 392
Dh 1674 ACCTTGGATTCCAGTACGAGCCCGGGTTCGAGAGCTCCAGAGAGTCCGACCCAAACC 1733
Qy 392 eSerValAspThrProProArgLeuSerThrProGluLysGlyProSerThrHisProL 412
Dh 1734 GAGCAGTTCAG--CTCTCTGGCAGACAACTCCAGCTTCCGCTGACCGGCGCC 1790
Qy 412 ysgLysAlaPheSerSerGluLeuLysAspLeuProTyLysSerThrThrGluMet 432
Dh 1791 TTGGATCTGCT-----C 1802
Qy 432 yrlLeuCys-----ArgTrpHisGln-ProProProSer----- 442
Dh 1803 ACATCTGCTCGACTGGAGGAGCTGGAGACTGCCACACCCCTCTGCTGCTCC 1862
Qy 443 ---ProLeuProLeuArgGluSerSerProLysLysGluGlu-----ThrVal 457
Dh 1863 ATTCCCGAGCTTCCGCGGGGCTGGCCCAAAACCTTGACCTCCGCGGCTCCCT 1922
Qy 458 AlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValProSer 477
Dh 1923 GGGGCGCTGCACCCCAAGTCTTCGAGCCTGCCCTGCACATCTGAGGGTCTCCGCA 1982
Qy 478 TrpArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuAsn 497
Dh 1983 AGCAGCAGCAAGAGGAGAGTCTCTCCAACTCCCTCCAGTCCGAGTCCATCCACAC 2042

RESULT 9
AAK52942
ID AAK52942 standard; cDNA: 4037 BP.
XX AAK52942;
AC AAK52942;
DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 2471.
DE Human, cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
OS Homo sapiens.
PN WO200157190-A2.
XX 09-AUG-2001.
PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0624936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HISE-) HISEO INC.
PA Tang YT, Liu C, Drmanac RT, Yasundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR WPT: 2001-476283/51.
DR P-PSDB: AAM79809.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PS Claim 1: Page 4748-4749; 622pp: English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 4037 BP; 1173 A; 1058 C; 1233 G; 573 T; 0 other:
Alignment Scores:
Pred. No.: 3.18 Length: 4037
Score: 235.50 Matches: 183
Percent Similarity: 32.05% Conserved: 83
Best Local Similarity: 22.05% Mismatches: 291
Query Match: 7.27% Indels: 273
DB: Gaps: 22
US-10-054-935-2 (1-614) x AAK52942 (1-4037)
Qy 11 AlaAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyArgAlaAlaAla 30
Dh 638 GCTGCGGAGCAGCAGAGGCGG-----CCGCTCCGCTATGGGCGAGCTGACGA 685
Qy 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuPArgHis 49
Dh 686 GCGGAGGTCCGCGAGATGCCGCGCGGTCTGCTGCTGGG-----CCGCGC 733
Qy 50 ArgLysLeuLysGluPro---GlyProProLeuAlaSerSerGlnGlySerPro 67
Dh 734 GCGGCTACGCTACGCTGAGCAGAGACCTGCTGAGACATCCGCGCAGTGGCGCA 793
Qy 68 AlaPro-----SerProAlaGlyCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
Dh 794 GCGCTAGACGACGAGGCGCGGAGCAGAGAGGAGGAGGCGGCGCGCGCTGGC 853
Qy 85 Ala-----GlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 97
Dh 854 GCGCTGCGCAGCAGAGCGCGGCGCGCGCGCTGACCTGACGAGAAGAGCGCGCGCT 913
Qy 98 ---GlySerValProLeuProCysProProAlaThrLysGlnAlaGlyIleGlyGly 116
Dh 914 GCAGGAGAGTGGCGGTACCTGCGCGCGCACCA----- 946
Qy 117 GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyArgAlaValLeuPro 136
Dh 947 ---CCAGCAGAGGTGGCGGAGCTGCTGCGCAATCCAGGCTCCGG----- 991
Qy 137 IleGlnThrGlySerLeuValAlaAlaAlaLysGluProThrProTPAlaGlyAspLys 156
Dh 992 ---GCGCGCGCAGCGCAGATGACGCGCGAGACCGCGCGCGCTGAAGTCCGACGTGAC 1048
Qy 157 GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro 176
Dh 1049 GTCGCGCTGCGCAGATTCGCGCGAGCTTGAAGCGCACGCGGTGACAGACGCTGCA 1108

```

QY 177 LeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
Db 1109 GTCCGAGGA-----GTGGTTCCGAGTGGAGGCTGGACCGACGTGCGAGGAGCCAA 1159
QY 197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGly----- 211
Db 1160 GGT-----GAAACACAGACGCTATCGCTCAGCCGACGAGAGGATRACTGAGTACCG 1210
QY 211 ----- 211
Db 1211 GCGTCAGCTGAGGCCAGACACAGAGCTGAGGCGACTGAAAAGCACCAGAGCACTACT 1270
QY 212 -----GlySerGlyAlaSerSerGlnAlaAlaCysLeu-LysGlnIle 225
Db 1271 GGAGAGCGACGGCTCTGAGCTGAGAGACCGTCATCAGGCCAGACATTCCTCCCTACAGGA 1330
QY 225 eLeuLeuLeuGlnLeuAlaPleu-----LleGluGlnGlnGlnGlnLeuGlnAlaLysGlu 244
Db 1331 AGCCATTCAGCAGCTGGAGCGCTGAGCTGAGACACACCAAGTGGAGATGGCCGCCAGCT 1390
QY 244 uLysGluIleGluLeu----- 250
Db 1391 GCGAAGATACCGAGACCTGCTCAATGTCAGATGGCTGTGATATAGATAGCCGCTTA 1450
QY 250 ----- 250
Db 1451 CAGAAACTCTCGAAGGTGAAGAGTGTGGATTGGCTTTGCCCAATTCCTTCTGCGCT 1510
QY 250 ----- 250
Db 1511 TCAGAAAGACTCCCAAAATTCCTCTGTGTCCACATCAATAAGGTGAAGAGGAGGA 1570
QY 251 -----LysSerGluArgAspThrLeuLeuAlaArgIleGluArgme 264
Db 1571 GAAGATCAAAAGTGTGAGAGTGTGAGAAAGAACTGTGAT-----GTGGAGGAGACA 1624
QY 264 tGluArgArgMetGlnLeuValLys-----LysAspAsnGluLys 277
Db 1625 GACGAGAGAGACCCAAAGTGAAGTGAAGAGTACTGAAAGAGAGATTAAGAGGCCAAGA 1684
QY 277 sGluArgHisLysLeuPheGlnGlyTrpGluThrGluLysArgIle----- 292
Db 1685 GAGAGAGGCGCAAGAGAGAGAGAGGGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1744
QY 293 -----GluThrGluLeuSerGluLysIle 300
Db 1745 AACAAAGTCTCCCCAGACAGAGAGGCTGCATCCCCAGAGAGAGAGAGAGAGAGAGAG 1804
QY 300 eLysLeuGluCysGln-----ProGluLeuSerGluThrSerGlu 313
Db 1805 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1864
QY 313 nThrLeu----- 315
Db 1865 CCCAGCCGAGAGTCAAGTCCCTGAGAGAGCCAAAGTCTCCAGCAAGAGAGAGAGAGAG 1924
QY 316 -ProPro-LysProPheSerCysLysArgSerGlyLysGlyHis----- 329
Db 1925 ACCGCTGAGAGGCCAAAGTCCCAAGAGAGATGAAAAGAAATTTCCAAAGCTGAGGTCA 1984
QY 330 -----LysArgLysSerProPheGlySerThrGluArgLysThrProValLysLys 346
Db 1985 AGTCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2041
QY 346 yLeuAlaProGluPheSerLys-----ValLysThrLysThrProLys 360
Db 2042 CCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2101
QY 360 yHisSerProIleLysGluGluProCysLys-SerLeuSerGluThrValCysLysArg 379
Db 2102 CCAAGTCCCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161

```

```

QY 379 ----- 379
Db 2162 CCAGTCTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2221
QY 380 -----GluLeuArgSerGlnIleuThrProGluLysProArgSerVal--- 394
Db 2222 GCCAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2281
QY 395 -----AspThrProArgLysSerThrProGluLysGlyProSerThrHisPro 411
Db 2282 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2341
QY 412 -----LysGluLysAlaPheSerSerGluIleGluAspLeuPro 424
Db 2342 GAGAGAGCCAGAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2401
QY 425 TyrLeuSerThrThrGluMetCysLysArgTrpHisGlnProProProSerProLeu 444
Db 2402 GTGAAGGAGAGAGAGAGAG-----TCCCTGAGAGAGAGAGAGAGAGAGAGAG 2440
QY 445 ProLeuArgGlu-----SerSerProLysLysGluIleuThr 456
Db 2441 CCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2500
QY 457 ValAlaArgCysLeuMetProSerSerValAlaGlyIleuThrSerValLeuAlaValPro 476
Db 2501 AAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2560
QY 477 SerTrpArgAspHisSerValGluProLeuArgAspProAsn-----ProSerAsp 493
Db 2561 GAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2620
QY 494 LeuLeuGluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLysGluLeuAsp 513
Db 2621 GCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2680
QY 514 GluLysArgArg-----LysArgTrpAspIleGluArg---Ile 525
Db 2681 GAGAGATGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740
QY 526 ArgGluGluArgGlyLeuGlnArgLeuGlnLeuArgTrpMetLysLysLysGlyIleGln 545
Db 2741 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2800
QY 546 GluSerGluProGluValThrSerPheProGluProAspAsp-----ValGluSer 563
Db 2801 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2860
QY 564 LeuMetIleThrProPheLeuProValVal-AlaPheGlyArgProLeuProLysLeuThr 583
Db 2861 CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2920
QY 583 rProGluAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeuGluIleGlu 603
Db 2921 CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2968
QY 603 nLysLysGluThrProHisArg 610
Db 2969 AAAAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

RESULT 10

AAD33243 standard; cDNA: 3608 BP.

AAD33243;

01-JUL-2002 (first entry)

Human secreted protein-encoding gene 7 cDNA clone HMWBp38, SEQ ID NO:17.
 Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
 immune disease; autoimmune disease; anemia; multiple sclerosis; cancer;
 rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;

KW	seazary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
KW	Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
KM	cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
KV	thrombosis; ocular disorder; corneal infection; wound healing; cardiant;
KW	vascular; thrombolytic; cytostatic; nootropic; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	97..3018
FT	/tag- a
FT	/product= "Human secreted protein precursor"
FT	sig_peptide
FT	97..198
FT	/tag= b
FT	mat_peptide
FT	199..3015
FT	/tag- c
FT	/product= "Human mature secreted protein"
XX	
PN	WO200218435-A1.
XX	
PD	07-MAR-2002.
XX	
PE	17-JAN-2001; 2001WO-US01567.
XX	
PR	28-AUG-2000; 2000US-228084P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
P1	Rosen CA, Komatsu GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
P1	Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M,
P1	Ni J;
XX	
DR	WPI: 2002-281060/32.
XX	
DR	P-PDSB: AAE20799.
XX	
PT	Isolated nucleic acid molecule encoding a human secreted protein is
XX	used in preventing, treating or ameliorating e.g. Alzheimer's disease,
PT	cardio-/cerebrovascular disorders and multiple sclerosis -
XX	
PS	Claim 1; Page 420-421; 504pp: English.
XX	
CC	AAD33337-AAD33280 represent cDNAs corresponding to 18 human secreted
CC	protein genes, and AAE20793-AAE20836 represent the proteins they encode.
CC	AAE20837-AAE20847 represent human secreted protein fragments. The genes
CC	and their corresponding secreted proteins are useful for preventing,
CC	treating or ameliorating medical conditions, e.g., by protein or gene
CC	therapy. Pathological conditions can be diagnosed by determining the
CC	amount of the new protein in a sample or by determining the presence of
CC	mutations in the new genes. Specific uses are described for each of the
CC	18 genes, based on the tissues in which they are most highly expressed,
CC	and include developing products for the diagnosis or treatment of
CC	immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
CC	Infections, anaemia, rheumatoid arthritis and multiple sclerosis),
CC	cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC	the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC	diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
CC	Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
CC	cardiac arrest, tachycardia, angina and thrombosis), infections caused
CC	by bacteria, viruses and fungi and ocular disorders (e.g. corneal
CC	infections). Secreted proteins of the invention can also be used to
CC	promote wound healing, maintain organs before transplantation, support
CC	cell culture of primary tissues, modulate differentiation of embryonic
CC	stem cells, induce mesodermal tissue to differentiate in embryos,
CC	modulate mammalian characteristics (e.g. height and weight), modulate
CC	the metabolism, anabolism, energy storage, mental state, biorhythms,
CC	cardiac rhythms, reproductive potential, hormonal levels appetite,
CC	memory and stress. They can also be used as an additive to increase
CC	or decrease storage capabilities and nutritional content of food. The
CC	present sequence represents a human secreted protein-encoding CDNA of
CC	the invention.
XX	
SO	Sequence 3608 BP; 738 A; 1077 C; 1199 G; 594 T; 0 other;

[illegible]

```

OY 278 luarGHisLysLeuPheGlnGlyTyrGluThrGlnGluArgGluuThrGluLeuSerg 298
Db 1889 ATAAACCCAGG-----CAAGTAGAGTGCATCAAAAGCCCGGAGCCG 1933
OY 298 luLyslleLysLeuGluCys---GlnProGluLeuSergGluThrSerglnThrLeuPro 317
Db 1934 TCCCCAGAGCCAGAGGCCCTGTAAAGCCAGAGCCAGAAACGGTGAATCTGCCTC 1993
OY 317 rGLysProPheSergSergLysArgSergLysGlyHisLysArgLysSergProPheGly 337
Db 1994 CCTGCTCTTGAC-----CCTGCTCTAGAGACTCTGGGCGC- 2031
OY 337 erThrGluArgLysThrProValLysLysLysLeuAlaProGluPheSergLysValLysThrL 357
Db 2032 -----CGCCCTGCTTCCATCCAGAGACTT-----AACGAGC 2062
OY 357 ystThrProLysHisSer-----ProLleLysGlnGluProCysGlySergLeuSergLut 375
Db 2063 GCTCCCTGGAGCACTGTGAGGGGCGCTGTGGGAGAGACCTGCTGCG- 2109
OY 375 hrValCysLysArgGluLeuArgSergLutThrProGluLysProArgSergSergValA 395
Db 2109 ----- 2109
OY 395 sphThrProProArgLeuSergThrProGlnLysGlyProSergThrHisProLysGlu---- 413
Db 2110 -----CCTGCTGAGGGCGCGCCCTGACACAGAGCTCGGAGAGCC 2149
OY 414 -----LysAlaPheSergSergGluThrGlnLysPheProTyrLysSergThrThrGluMet 432
Db 2150 AGGCGAAGCGAGAGATGGCCAGAGATGCCGCCCGCCAGAGGAGCTGCACTGTGAAG 2209
OY 432 yrlLeuCysArgTyrPheGlnProProSergProLeuArgGluSergSergProL 452
Db 2210 AGCTCCCAAGGCGCCGAGAGAGTGCCTGCCAGACCCGCGGAGAGAGCGGGGCGC 2269
OY 452 yslYsglu 454
Db 2270 CAGAGAG 2277

RESULT 11
AA158822
ID AA158822 standard; cDNA; 3169 BP.
XX
AC AA158822;
XX
DE 22-OCT-2001 (first entry)
XX
DE Human polynucleotide seq ID NO 1025.
XX
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.

```

```

PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX
DR P-PSDB: AAM39666.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 1025; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neotropic.
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 3169 BP: 697 A; 922 C; 927 G; 623 T; 0 other;
XX
XX
Alignment Scores:
Pred. No.: 4.6 Length: 3169
Score: 228.50 Matches: 151
Percent Similarity: 36.25 Conservative: 77
Best Local Similarity: 24.018 Mismatches: 199
Query Match: 7.068 Indels: 204
DB: 22 Gaps: 32
XX
US-10-054-935-2 (1-614) x AA158822 (1-3169)
OY 13 AlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
Db 47 AGTCCCGCTCCCGGCGTCT----- 67
OY 33 GlyProGluAspGluProGlyAlaAlaGlnAlaHisPheLeuProAlaGHisArgLysLeu 52
Db 68 TCACACATGATGCCACCGCGCCGAGAGCCCGCCGCGCCGATGCCGTGGCC 127
OY 53 LysGluProGlyProProLeuAlaSerSergGlnGlyLysSergProAlaProSergProAla 72
Db 128 AACAAGAGGCG-----TCCAGGCGCGCCAGCGCGCGGAGCCCTCG 172
OY 73 GlyCysGly-----GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro 89
Db 173 GAGAAAGGTGCCACCGCTGCGCGCGCGCGCTGCGCGCTGCGCGAGCGCAGCGCGCC 232
OY 90 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 283
Db 233 GCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 283
OY 110 LysGlnAlaGlyLleGlyGlyGlnProAlaAlaAlaGlyAlaGlyCysSergProArgPro 129
Db 284 GCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
OY 130 LysTyrGlnAlaValLeuProIleGlnThrGlySergLeuValAlaAlaAlaLysGluPro 149
Db 344 -----TCGGCCAAACAA--- 355

```


QY 150 ThrProTrrAlaGlyAspLysGly-----GlyAlaAlaSerProAlaAlaThrAlaSer 167
 Db 356 -----AGGGCTCCAGAGGGGGCGACGGCGCGGAGCCCTCGAGAAAGGTGCC 406
 QY 168 AspProAlaGly-----ProProProLeuProLeuProGly 179
 Db 407 CACCGCTGCGGGCGGGGATGACGTGGCGAAGAAGACCGCGCGCGCCACAGCCG 466
 QY 180 ProProProLeuAlaPro-ThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTr 199
 Db 467 CCGCGCGCGCGCGCGCGCGACCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
 QY 199 PLYSerMetArgLysSerProLeuGlyGlyGlyGlySerGly----- 214
 Db 524 GCGGACGCGCAAGGGCGGCAACCGC-GGCGGGGGGGCGCGCGCGCAAGTCTCTCTC 582
 QY 215 -----AlaSerSerGluAlaAlaCysLeu 223
 Db 583 CTCTCCGCGCTCCG 642
 QY 223 sGlnIle-----LeuLeuLeuGlnLeuAspLeuIle----- 233
 Db 643 CAGGCTGCGCAGGGCGCTCAACTTCTCTTCTTCTTCTGCGCGCGCGCGCGCG 702
 QY 233 ----- 233
 Db 703 CTGCGGGTGGTGGCTCCACACGTCCTGAGAGAGTCCAGAGGTCCGCGCGCACCA 762
 QY 234 -----GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
 Db 763 GGAATTCTCCGCGAG 822
 QY 247 eGluGluLeuLysSerGluArgAspThrLeuAlaArgIleGluArgMetGluArg 267
 Db 823 GCACTCTTGCAACCCACATTGGAACCTTTGAGTCCACTCTTGAGAAAGTCCCAACAT 882
 QY 267 gMetGlnLeuValLysLys-----AspAsnGluLysGluArg---HisLys 281
 Db 883 ACAAGACTCTCAGAGAAAGCTGTGAAGCAGAGGAGAGAGTGAAGTCAAGCGGATCAG 942
 QY 281 sLeuPheGlnGlyTrpGluGlu----- 289
 Db 943 AGTCTGCGAAGACTCCAGATGAGATTCCTCAAGACCTCTCGATGGATCCATGTG 1002
 QY 290 -----GluArgGluGluThrGluLeuSerGluLysIleLysLeuGluGly 305
 Db 1003 GAAAGAGCCCGCGGAGGGGAGCTTCAGCTCCCTGAGAACGCGGAG---GAGCGGCT 1059
 QY 305 nProGluLeuSerGluThr-SerGlnThrLeuProPolysProPheSerCysGlyArg 325
 Db 1060 GAGGAGAGCTCCAAATCCATCAACGACACATCGCCA-----TCTTCACAGAGT 1110
 QY 325 eArgLysGlyHisLysArgLysSer-----ProPhe----- 335
 Db 1111 CCAGAGAGGAGCCAGAGAGATCAATGACATGAAGCAAGGTTGCCCTCCGAGAGA 1170
 QY 336 -----GlySerThrGluArgLysThrProValLysLeu----- 347
 Db 1171 ATCTGAGGGAGCAACAGCATTTGAAGACCTTAAAGAGACCTGTGAAGAGATACAGAC 1230
 QY 348 -----AlaProGluPheSerLysValLys----- 355
 Db 1231 CTCAGCCCAAGTCAGAGAGTGGACATGAGAGCCCTGAGAGTACCTTCAGACATGGA 1290
 QY 356 -----ThrLysThrProLysHisSer----- 362
 Db 1291 GTCTGACATCTACACGAGGTCCGAGCTGTGAGCCTCAAGCAGAGCAGAGGCTTT 1350
 QY 363 -----ProIleLysGlu---GluProCysGlySerLeuSerGluThrValCysL 378
 Db 1351 CAAGAGAGCGCGCCACACAGAGAGCGGCTGCGCTGACG-CGCTTCACGAGAGAGCTTCA 1409
 QY 378 ysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProp 398

Db 1410 GGTCTGAGAGAGTCCGTCTCCCGCTCCCGGAGAGAGATCCGAGACTGGAGAGAGCTCC 1469
 QY 398 roArgLeuSerThrProGlnLysGlyProSer-----ThrHisProLysG 413
 Db 1470 GCCAGCTGAAGTCCGATTCACACGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1529
 QY 413 Lu---LysAlaPhe-SerSerGluIleGluAsp---LeuProTyrlLeuSerThrThrGlu 430
 Db 1530 CTTTGAGGCACTCCAGCAAAAGATCAGGAGACTGAGACTCCAGGCTCCAGACGTGAGG 1589
 QY 431 MetTyrlLeu-----CysArgTrp-----HisGlnPro-ProProSerProLeuPr 445
 Db 1590 ATGGGGTGCTCTCCATGATGAGGTGCTTCTGCGCGCAGACAGCAGAGCCTGAGTCCCTCC 1649
 QY 445 OleuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSe 465
 Db 1650 TGTCAGAGAG-----CCAGAGACAGCAGAGCGCGCTGCGCGC-----CCTGCAGG 1694
 QY 465 rValAlaGlyLutThrSerValLeu---AlaValProSerTrp---ArgAspHisSerVa 483
 Db 1695 GCGCGCTGAGAGCGCTCGGCTCTCAGAGCAGACAGAGATGCGCTGCGCACAGCGTGA 1754
 QY 483 IgLuProLeuArgAspPro 489
 Db 1755 GGAGCGTGGCGGAGACCA 1773
 RESULT 12
 AA160608/c
 ID AA160608 standard; cDNA: 3203 BP.
 AC AA160608;
 DT 22-OCT-2001 (first entry)
 DE Human polynucleotide seq ID NO 4597.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 KW OS
 KW Homo sapiens.
 PN WO20015312-A1.
 PD 26-JUL-2001.
 PF 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HXXSE-) HXXSEQ INC.
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
 PI Zhao QJ, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM41452.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

PS Claim 1: SEQ ID NO 4597; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA158642-AA162213) with neotropic
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral neuropathies, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilization of the activities such as: immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, hemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 3203 BP; 629 A; 940 C; 931 G; 703 T; 0 other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
Percent Similarity:	4.64	3203	151
Best Local Similarity:	228.50	Conservative:	77
Query Match:	36.25%	Mismatches:	199
	24.01%	Indels:	204
	7.06%	Gaps:	32

US-10-054-935-2 (1-614) x AA160608 (1-3203)

```

OY 13 ALAPROALAGLYLALYSPROGLUGLARGLEUASPTRYGLARGALALALALEUGLY 32
DB 3157 AGTCCCGCTCCCGGGGCTCCT-----
OY 33 GLYPROGLUASPGLUPROGLIYALALAGLUALAHSPHELEUPROARGHISARGLYLEU 52
DB 3136 TCACACATGATGATCCAGCCCGCCGAGCCCGCCCGCCGATCCCTCCGCC 3077
OY 53 LYSGLUPROGLYPROPROLEUALASERSEGLNGLYSEPRALAPROSERPROALA 72
DB 3076 AAMCAAGAGGCG-----TCCAAAGCGCGGCCAGCGCCGCCGCCCTCG 3032
OY 73 GLYCYSGLY-----GLYLYSGLYARGLYLEULEUPROALAGLYALAALAPRO 89
DB 3031 GAGAAAGGTGCCACCCGCTCGCGCGCTCTCTCGCGCTCCGACGCCAGAGCGCC 2972
OY 90 GLYGLNGLUGLUGLUSERTTPGLYGLYSEVALPROLEUPROCTYSPROPROALATHR 109
DB 2971 GCGTGCCTGCGCCAGTCCCGGGGAGCC-----CTGCAAGTTCCCGGGCCG 2921
OY 110 LYSGLNALAGLYLEGLYGLYGLUPROALAALALAGLYALAGLYCYSSERPROARPRO 129
DB 2920 GCGCGCGCTCGCTCGCTCCAGCCGCGGCCGAGCCGCCGCCCATGCC 2861
OY 130 LYSTYGLNALALVALLEUPROILEGLNTHRGLYSERLEUVALALALALYSGLU 149
DB 2860 -----TCCGCCAAACAA---
OY 150 THRPROTRALAGLYASPLYSGLY-----GLYALALASERPROALALATHRALASER 167
DB 2848 -----AGGGGCTCCAAAGGGCGCCAGCGCCGCGGAGCCCTCGGAGAGGGTCC 2798
OY 168 ASPPROALAGLY-----PROPROLEUPROLEUPROGLY 179
DB 2797 CACCGGTGCGGGCGCGGATGACGTGCGAGCAAGAACGCGCGCGCGCGCGAGCGCG 2738
OY 180 PROPROPROLEUALAPRO-THRALATHRALAGLYTHLEUALALASERGLUCLYARGTR 199
DB 2737 CCGCGCGCGCGCGCGCGAGCCGACAGCCAGC-----ACCGCAGAGAGCAGCCAGAAC 2681
OY 199 PLYSERMETARGLYSERPROLEUGLYGLYGLYGLYSEGLY-----
DB 199 -----

```

```

DB 2680 GCGCAGCGCAAGGCGGCCACCGC-GCGCGCGCGCGCGCGCGCAAGTCTCTCTC 2622
OY 215 -----
DB 2621 CTCCTCCGCCCTCCGCCGCGTGCAGCGCGCGCGCGCGCGCTCTCTCTCTC 2562
OY 223 sglnlle-----LeuleuLeuLeuAspLeule- 2502
DB 2561 CAGGCTCGGCAAGGCGCTCAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2502
OY 233 -----
DB 2501 CTCGCGCTGCTGCTCCACACGCTCGAGAGAGTCCAGAGCTCCGCGCGAGCCACA 2442
OY 234 -----
DB 2441 GCACTCTCCCGCGAGAGGAGAGCTGGCGCCAGGCTTCGAGCGCTCGAGCAAGAGT 2382
OY 247 eglugluLeuLysSerleuArgAspThrleuLeuAlaArgIleuArgMetleuArg 267
DB 2381 GCACTCTTCGAGCGCCACATTTGACCTTTGAGTCCATTTGAGAGCTCCCAACATA 2322
OY 267 gmetgluLeuValLysLys-----AspAsnLysLysLysLysLysLysLys 281
DB 2321 ACAAGACCTCACAGAGAAAGCTGTGACAGCAAGGAGAGTGTGAGCGGATCAGCCA 2262
OY 281 sleupheuglyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 289
DB 2261 AGTGCCTGAGAACTCCAGAAATGAGATCTCAAGAGCTCTCGATGAGATCCATGTGT 2202
OY 290 -----
DB 2201 GAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2145
OY 305 nProgluLeuSerleuThr-SerleuThrleuProProLysProPheSerleuArg 325
DB 2144 GAGGAGCTCACAAATTCATCAACAGCAACATCATGCCA-----TCTTCAAGAAAGT 2094
OY 325 eRgLYGLYGLYHISLysArgLysSer-----ProPhe----- 335
DB 2093 CCAGAGAGAGAGCGCAGAGAGATGATGATGAGAGCAAGGTTGCCCTCTCGAAGA 2034
OY 336 -----
DB 2033 ATCTGAGGGAACAAACAGAGATTGAAGCTTTAAGAGAACTGTGAAGAGATACAG 1974
OY 348 -----AlaprogluPheSerLysValLys----- 355
DB 1973 CTGAGCAAGTCCAGAGAGTGGACATGAGAGCGCCCTGAGAACTACCTTACATGGA 1914
OY 356 -----ThrlYsThrProLysHisSer----- 362
DB 1913 GTCTACATCTACACCGAGTCCCGAGCTGTGAGCTTCACAGAGAGAGAGCTTT 1854
OY 363 -----ProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1854
DB 1853 CAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1795
OY 378 ysarggluLeuArgSerleuLysLysLysLysLysLysLysLysLysLysLysLys 1795
DB 1794 GGTCTGAGAGAGTCCGCTCCCGCGCTCCCGAGAGATCCGAGATCGAGAGAGAGCTCC 1735
OY 398 roArgleuSerThrProgluLysLysLysLysLysLysLysLysLysLysLysLys 1735
DB 1734 GCGAGCTGAGAGTCCGATCCACAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1675
OY 413 lu-----LysAlaPhe-SerSerleuLysLysLysLysLysLysLysLysLys 1675
DB 1674 CTTTGAAGAGCTCCAGCAAAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAG 1430
OY 431 MetTyrLeu-----CysArgTyr-----HisGlnPro-ProProSerProleu 445
DB 1614 ATGGGCTGCTTCATGAGAGTGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1555

```


Pred. No.:	6.78	Length:	3117
Score:	223.50	Matches:	185
Percent Similarity:	31.67%	Conservative:	76
Best Local Similarity:	22.45%	Mismatches:	210
Query Match:	6.90%	Indels:	353
DB:	21	Gaps:	42

US-10-054-935-2 (1-614) x AA238858 (1-3117)

Oy	12	AlaIalaProAlaIaGlyI	-----	AsnProIuGlnIaIaGlyAspTyrGluIaArgAla	--- 28	
Db	293	GCTGTCCAGCAGAGGTGGCTCAGTAAAGTGGCCCACTCTGTGATGGCTCCAGGGCGGGG			352	
Oy	29	-----	-AlaIalaLeuGlyI	Pro	34	
Db	353	CCCGAGAGGTGGAGAGAGTGGGGGTGCATGACAGCAGCTGGCAGCAGCAGCCAGGCCA			412	
Oy	35	GLuAspGluProGlyAlaIaGluAlaHisPheLeuProArgHisArgLysLeuLys	--- 53			
Db	413	GCTTCACACCATTCAACATGAA	-----	CTGCACAGCTGTCTCCCTCAACAGTGC	463	
Oy	54	-----	-GluProGly	-----	ProIleuAla	60
Db	464	AACCAAGGGAGCCCGCATCTCGAAGTGGAGTCCGACTACTATTAACACCCCTGAGGC			523	
Oy	61	SerSerGlnGly	-----	GlySerProAlaProSerPro	----- 71	
		:::				
Db	524	ACTGAAGGGAGAAAGGGGGGGCCCACTGACCTGACCGCTATGTGCACCAATGATGCC			583	
Oy	72	-----	-AlaGlyCysGlyGlyLysGlyLysGlyLysArgGlyLeuLeuLeuProAlaGlyAlaAlaPro		89	
			:::			
Db	584	ACAGAAGGTGCAGCTGAGGTAGC	-----	-----	GGGGCCCA	616
Oy	90	GlyGlnGlnGlu	-----	GluSerTyrGlyLysSerValProLeuProGlySerProProProAla		108
Db	617	GGCACCCCTGAAATTCCTTCCACGACGACCCAGAAACC	-----	CCCAACAGTCACGTCCCTGCA		676
Oy	109	ThrySglnAlaGlyIle	-----	GlyGlyGluProAlaAlaAlaGlyAlaGly	--- Cys	125
Db	677	ACCCTTCACGTGGCATTCGGCCACACAGAGTAAACCGGAGGCTTCCTCCGGCAGGGCCACC				736
Oy	126	SerProArgPro	-----	-LysTyrGlnAlaValLeuProIleGlnThr		139
		:::				
Db	737	GCCCCCAACCCGGGTGCTTCCTCCACAGAGAGAGAGCAGCAGCAGCAACACCA				796
Oy	140	GlySerIleValAlaAlaAlaLysLysLeuProThrPro	-----	-----		151
		:::				
Db	797	GCAGCAGCAGCAGCAGCAGCAGCAGCAGCCTACCCAGATGCCGCTTTTGAAACTTCTA				856
Oy	152	-----	-TyrAla			153
Db	857	TTCCATGCCAGCAGCAACCTCGCAGCAACCCAGACCTTGTGGCTGCAGCAGCTGGGCC				916
Oy	153	-----	-----	-----		153
Db	917	ACTGGAGACGTCCACACTGTGCTCACACAGCATGGCAGCACCCTTCCCCCCCACACC				976
Oy	154	-----	-GlyAspLysGlyGlyAlaAlaSerProAlaAlaThr			165
						:::
Db	977	AGATATGAACCCAGAACTGCCAAGCCCTTCTGCGAGA	-----	CTCAGGCCCGCAGCCACGC		1033
Oy	166	AlaSerAspProAlaGlyProProProLeuProLeuProGlyProProProLeu	-----	-----		183
Db	1034	GCTTACC	-----	TCAAGTCCAGATCCCTTCCCGCGGCTCCGCGCCCTCTCTAAGA		1087
Oy	183	-----	-----	-----		183
Db	1088	GGGATATCTGCTCCACAGCCCTGATGGGCTGGCAGCCAGCTGGCAGAGAGGCCAC				1147
Oy	183	-----	-----	-----		183
Db	1148	TGGCAACCTGTCTTCAATCACTAGCCCTCGACAGCCGACCCACCTGGCTCCCTGGGCA				1207

QY	184	ALAProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGly-----	197
DB	1208	GGCCATCATCTTAAGAGCTGGAGATTCCCGCTGGAGGTGAAGGATGCACTGCTGA	1267
QY	198	-----ArgTrpLysSerMetArgLysSerProLeuGlyGlyGly	211
DB	1268	TGGGAGAGACTAGACCAACCAATGGCCGGGAGCGAGAGAGCTCTGGCATGGCAGGAGGA	1322
QY	212	GlySerGlyAlaSer-SerGlnAlaAlaCysLeuLysGlnIleLeuLeuGlnLeuAs	231
DB	1328	GGGCATGAGGGCAGATGAGACACAGGGAGCTGTGGCGAG-----GTGCTACAGGGCGG	1378
QY	231	PLeuIleGluGlnGlnGlnGlnGlnGlnAlaLysGluLysGluIleGluGluLeuL	251
DB	1379	AGTGTATCCAGAGACACGACGGAGCGGCGCGGCA-----	1412
QY	251	SerAlaLysArgAspThrLeuLeuAlaArgIleGluLysMetGluArgArgMetGlnLeu	271
DB	1413	-TCCAGGAGGCCAATTGCTGACC-----CTGGCCGAAAGGCTGTGGAGCTGGC	1462
QY	271	LysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTrpGluThrGluLys	291
DB	1463	CTCACTGCAGAGATGCAAGAGAT-----GGCAGTGGTTCGAAAGAGA	1504
QY	291	GluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGluTh	311
DB	1505	GGCGAAAGATGATGATGGCTCAACTACCAAG-----TGTGGGTGTGAGTTTCTGAGGC	1558
QY	311	rSerLysThrLeu-ProPolys-----ProPheSer- 321	
DB	1559	TTCC-----TTAGCCACCAAGCGAGACGAGACAGACAGATGGGATGTATCCCTCATCT	1612
QY	322	-----CysGly-----	323
DB	1613	CCCACTGTCTGTGGCTGTGGCAACTGTGACACCACTAGAGCGCAGCGCGTGAAGTCT	1677
QY	324	----ArgSerGlyLys-----GlyHisArgLys 333	
DB	1673	TGATGAGAGCGGAGAGGCTGTGACAGACACCCTGTCAGCAAGCCATCATGATCATCT	1733
QY	333	ePProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerL	353
DB	1733	CACCCGACGGCGGCTCCACCC--GATCCCGCGGACAGATGCTC-----A 177	
QY	353	ysValLysThrLysThrProLysHisLysSerProIleLysGlnGluProCysGlySerL	373
DB	1775	AGCTACGGCGGAGACATGAATGTCAAGTTGGAGGGGAGAGCTTCGTCGGGAACCAAA	1833
QY	373	eP-----GluThrValCysLysArgGluLeuArgSerGlnGluT 386	
DB	1835	CGAGGGCGCCAGGCCGAGCCCTCATCTATCCCAACGAAGGGGCGACTTTATCGCCCC	1894
QY	386	hrProGluLysProArgSer-SerValAspThrProProArgLeuSerThrProGlnL	405
DB	1895	TCCCGCTACTCCACATCATACCCATACAGAGCCACCTGCGCTCTCC--CCT 1944	
QY	406	GlyProSer-ThrHisProLysGluLysAlaPheSerSerGluIleGluAspProTy	425
DB	1946	GGGCTTAGTACACACCCCTCTGACGAGAGCTT-----GACTACTCTCC 1994	
QY	425	rLeuSerThrThrGluMetLysArgTrpHisGlnProProProSerProLeuP	445
DB	1991	C-----TACACGCGCGCCCCCATCTCAGGCC 201	
QY	445	oleuArgGluSer----- 450	
DB	2018	TGTGGGGGAAGGCTCTGGCTCTACTTCATATCCATCATATACACACAGCACCATCCCTGC	2077
QY	451	-----ProLysLysGluGluThrValAlaArgCysLeuMetProSer----- 464	
DB	2078	CCCTCTCTCCATACAGCGCTTAAGAGTGGCCATGTGACACGCTGCTCGGAGTAAACAGTGTGA	2133
QY	465	-----SerValAlaGlyGluThrSerValLeuAla----- 474	

```

Db 2138 AGTAAACCCCGCTCTCTCTGTGTGAGGGGAGGCCACCCAGTCAGCATCGACGACG 2197
OY 475 -----ValProSerTrpArgAspHisSerValGI 484
Db 2198 GATCAACGTGGGCTCCCGGTTCAGAGCAAAATCCCTTGATGAGGACCGTCCCTG-- 2255
OY 484 uProLeuArgAspProAsnProSerAspLeu-----GluAsnLeuAspAs 500
Db 2256 -GCAGCTGCAGATCCCAAGAGCTGACTGTGTGGCAGACCATGGAGAGCACTAGAGG 2314
OY 500 pSerValPheSerLyArgHisAlaLysLeuGluLeuAspGluLyArgArgLyAspTr 520
Db 2315 CAGC-----CGGAGAAACAGAGGCAAGTGA 2341
OY 520 pAsp----- 521
Db 2342 AGACCTGCTGACAGCCGCTGCTCCAGCATTTTCCCTGTGTGTGCACACACAGAGACT 2401
OY 522 ----1IeGIArgGLeArgGluGlnArg-----1IeLeuGlnArgLeuGlnLeuArgMe 538
Db 2402 GGCCCTGCACACTGTCTGCAGAAATCCAGAGAGAGACATCTGTGAAGCGTGAATAGCTGCT 2461
OY 538 tTrLyLysLyLysGly1IeGlnGlnSerGluProGluValPheSerPheProGluPr 558
Db 2462 GCTGAAGAAGCCCTGCGGCCCAACAACATCCGCTGCACACTTATCACTACACAGGCTC 2521
OY 558 oAsp 559
Db 2522 TGAC 2525

```

RESULT 15

AA578091
ID AA578091 standard; cDNA; 3414 BP.

AC AA578091;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #13895.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PDB; ABG13904.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT biotechnology for genetic disorders or other traits and to assess

XX Claim 1, SEQ ID NO 13895; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. AA564197-AA594564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pcr_sequences.

Sequence 3414 BP: 806 A; 1013 C; 838 G; 757 T; 0 other;

Alignment Scores:

Pred. No.:	7.3	Length:	3414
Score:	223.50	Matches:	164
Percent Similarity:	34.72%	Conservative:	87
Best Local Similarity:	22.68%	Mismatches:	256
Query Match:	6.90%	Indels:	217
DB:	23	Gaps:	37

US-10-054-935-2 (1-614) x AA578091 (1-3414)

```

OY 10 ALAALAAlaProAlaGlyLysn----- 18
Db 915 GCTGTGCTGCTCCCTGGAGTGGCTCATGTATGTAGAGTCACTTTGGAGATCTGC 974
OY 19 -----ProGlu-----GlnArgLeuAspLyGlyArg-AlaAla1 30
Db 975 CAGAACAAACATTCCTCGATTCCTAATAAAGCATGATGAGAGGAGACAGCT 1034
OY 30 ALeugLyGlyProGluAspGluPro-----GlyAlaAlaGluAlaHisPheLeuPr 47
Db 1035 GCTGTGCTGCTCCCTGGAGTGGCTCATGTATGTAGAGTCACTTTGGAGATCTGC 1094
OY 47 oArgHisArgLysLeuLysGlyLysProGly-Pro-----ProLeuA 60
Db 1095 ACTTTGGCGGACACTTGAGAGTCTTCGGCCACACATGCACATGAGAGGCCCTTTCT 1154
OY 60 lAsrSerGlnGlyLysSerProAlaProSerProAlaGly---CysGlyGlyLysGly- 78
Db 1155 GTTCTGGCCAGGCGCACAGCCGGCTCCTCAGCTTGCAGGATGTGTGAGGAGAGGCG 1214
OY 79 -----ArgGly-----LeuLeuLeuProA 85
Db 1215 CGAGCGGAGACCGGGGCTGTGCGGGCCCTTGCAGGCGCAGCTGGAAGCTTCTGTGGGTG 1274
OY 85 lAgLy---AlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 104
Db 1275 TGGGCTTGGCGGCGCCGCTCAGTCCGAGACGCGGCTG-----CCCT 1316
OY 104 ySPProPro-----AlaThrLysGlnAlaGlyTlAgly-----G 116
Db 1317 GCTG-GCTCCAGGCAATGAGGAGGACTTGAACCCGCGGCGGCTGTGAGGAGTACTG 1375
OY 116 lYgLuProAlaAlaAlaGly-----AlaGlyCysSer---ProArgProLyTyGlnA 133
Db 1376 GGTCCCGCAGAGTGGCCAGCCACCGAGTGTCTGATTTCTCGCGGGCCCTTACGTG 1435
OY 133 lAValLeuProLlGlnThrGlySerLeuValAlaAlaAlaLysGluProThrPro---- 151
Db 1436 CTTTCTGAGGAGGAGGCTGGGAGCTTGCAGCCCGCATGCTGAGGCTTCCCGGCT 1495
OY 152 -----TrpAlaGlyAspLyGlyLysAlaAlaSerProAlaAlaThrAlas 167
Db 1496 CCGTAACTTCTGTGAGCTGGAGCTCCCGCAGAGCGCGCCCTTCTCAGGCGCG 1555

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 22:10:06 ; Search time 279 Seconds

(without alignments)
3474.922 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTMRSAVFKAAPAGNPE.....NSRCLRIQKQTPRTCK 614

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1105431 seqs, 789497651 residues
Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_SPOOL/US10054935/runat_14072003_115026_24714/app.query.fasta_1.775
-DB=Published_Applications_NA -QEXT=fastcap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10054935 @CGN 1.1.57 @runat_14072003_115026_24714
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEJURY -NEG_SCORES=0 -WAIT -DSPBLCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1448.5	44.7	1529	10	US-09-925-302-212
2	1055.5	32.6	615	10	US-09-917-800A-725
3	625	19.3	365	9	US-09-918-995-30064
C	4	243.5	7.5	9025608	9 US-10-156-761-1

5	234.5	7.2	3107	10	US-09-783-066-5	Sequence 5, Appl1	
6	230.5	7.1	7065	9	US-09-991-496-115	Sequence 115, App	
7	230.5	7.1	7065	10	US-09-874-923-115	Sequence 115, App	
8	228.5	7.1	3169	9	US-10-037-270-713	Sequence 713, App	
9	226.5	7.0	13952	9	US-10-171-311-51	Sequence 51, Appl	
10	225	6.9	2108	10	US-09-962-832-225	Sequence 225, App	
11	223.5	6.9	3117	10	US-09-752-639-6	Sequence 6, Appl1	
12	223.5	6.9	3117	10	US-09-984-198-6	Sequence 6, Appl1	
13	222.5	6.9	8459	9	US-09-563-728A-31	Sequence 31, Appl	
14	221.5	6.8	6354	9	US-10-084-817-158	Sequence 158, App	
15	221.5	6.8	8459	10	US-09-817-913-8	Sequence 8, Appl1	
16	221.5	6.8	8459	10	US-09-817-538-8	Sequence 8, Appl1	
17	220	6.8	13555	9	US-09-764-891-7281	Sequence 7281, Ap	
18	219	6.8	9025608	9	US-10-156-761-1	Sequence 1, Appl1	
19	218	6.7	6457	10	US-09-880-107-3389	Sequence 3389, Ap	
20	217	6.7	36778	9	US-09-860-846-5	Sequence 5, Appl1	
21	217	6.7	36778	9	US-09-836-821-5	Sequence 5, Appl1	
22	217	6.7	36778	10	US-09-861-289-5	Sequence 5, Appl1	
23	217	6.7	37948	9	US-09-988-3848-5	Sequence 5, Appl1	
24	217	6.7	38506	9	US-09-793-708-19	Sequence 19, Appl	
25	216	6.7	33023	10	US-09-880-107-3350	Sequence 3350, Ap	
26	213.5	6.6	2725	9	US-10-177-293-117	Sequence 117, App	
27	213.5	6.6	2725	10	US-09-962-436-40	Sequence 40, Appl	
28	213.5	6.6	4437	9	US-10-156-761-4189	Sequence 4189, Ap	
29	212.5	6.6	2732	9	US-10-175-523-193	Sequence 193, App	
C	30	211.5	6.5	3331	9	US-09-373-658-31	Sequence 31, Appl
C	31	210	6.5	57130	10	US-09-835-081-3	Sequence 3, Appl1
32	208	6.4	14800	10	US-09-954-456-1601	Sequence 1601, App	
33	207.5	6.4	1855	10	US-09-864-761-19708	Sequence 19708, A	
34	207.5	6.4	1958	10	US-09-864-761-2927	Sequence 2927, Ap	
C	35	207.5	6.4	2001	9	US-10-156-761-3168	Sequence 3168, Ap
36	207.5	6.4	5001	9	US-10-156-761-2323	Sequence 2323, Ap	
37	206.5	6.4	2520	9	US-09-864-636A-573	Sequence 573, App	
38	206.5	6.4	2520	9	US-09-864-636A-577	Sequence 577, App	
C	39	206.5	6.4	3367	9	US-10-001-873-22	Sequence 22, Appl
40	206	6.4	2445	9	US-09-864-636A-571	Sequence 571, App	
41	206	6.4	2445	9	US-09-864-636A-575	Sequence 575, App	
42	206	6.4	2445	9	US-09-864-636A-579	Sequence 579, App	
43	206	6.4	2445	9	US-09-864-636A-581	Sequence 581, App	
44	206	6.4	2538	9	US-10-156-761-6189	Sequence 6189, Ap	
C	45	205.5	6.3	4826	10	US-09-772-304A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-925-302-212
; Sequence 212, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925, 302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-212

Alignment Scores:
Pred. No.: 1.05e-94
Score: 1448.50
Percent Similarity: 95.64%
Best Local Similarity: 95.30%
Query Match: 44.73%

Length: 1529
Matches: 284
Conservative: 1
Mismatch: 6
Indels: 7

OY	76	GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAla	88
Db	261	GGCGTGGGCAAGGCTGTGAACCGGACAGCTAGCCGTCCTCCAGGCTGTGCATT	320
OY	89	-----ProGlyGlnGlnGlnGlnSerTrpGlyGlySerValProLeuPro	103
Db	321	CAGACACCCCTTTGGCTGGACAGACAGAGCTGATGTCTCATGGAGA	371
OY	104	CysProProProAlaTrhTrpGlnAlaGlyIleGlyGlyGluProAlaAlaAlaGlyAla	123
Db	372	TGTCTCTCTCCGCTGTGTGAGCTCGAGACAGCTGG-----CCCGCGGCTCCGCGCC	425
OY	124	GlyLysSerProAlaGProLysTyGlnAlaValLeuProIle-----GlnThrGlySer	141
Db	426	GGCCA-GGACCCCAACCCAGCCACCCCGGAGCCTGTGACCTGACCTGCCACCACTCGGGAAG	484
OY	142	LeuValAlaAlaAlaAlaLysGluProThrProTrpTrpAlaGlyAspLysGlyAlaAlaSer	161
Db	485	CTTCTGCAGATATATCCACAGTCCGAGCC-----GGCTCTCCCAATGAGTGGGCGCCGACG	541
OY	162	ProAlaAlaTrhAlaSerAspProAlaGlyProProProLeuPro-----	178
Db	542	GCAGCTCTTACCTCCACATCCCTGAGGGGCGCTCCAAAGCAATCCTGTGCTCCCTCC	601
OY	179	-----GlyProProProLeuAlaAlaProThrAlaThrAlaGlyTrhLeuAlaAla	195
Db	602	TGCCGAGGAGGCTCTCCAAACCACTGTCCGACAGGCCCAACCAACACATGTATCAGTC	661
OY	195	rGluGlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlyGly-----	212
Db	662	CACACCCCA-----GACCCCTCGAAACCATCTCCTCTGTCTGTGGGAGAGCATCGAATC	718
OY	213	-----SerGlyAlaSerSerGlnAlaAlaCysLeu	223
Db	719	CGCCAGCCCATCCGAGGGGTCCCGAGAGGGGTGCTCTCGCCACAGCAGCACCACT	778
OY	223	ysGlnIleLeuLeuLeuSlnIleuAspLeuIleGlnGlnGlnGlnIleuGlnAla	243
Db	779	CCGAGGT-----AAATGACACAGCTTTGCAGACGGCAAGAAACAGGTGGTGAAG	829
OY	243	ysGluLysGluIle-----GluGluLeu	251
Db	830	CAGAGGCTGAGGCGGAGCGGGGTCCCCACAGTCCATCCGGCCCGCGGAGGCGCTGA	889
OY	251	ysSerGluArgAspTrhLeuLeuAlaArgIleGluArgMetGluArgMetGlnLeu	271
Db	890	GCTCGAGATGAGTGAAGCTCAGCGCGCGGTGAGAGAAACGACAGCATCGAGGCTC	949
OY	271	alLysLysAspAsnGlu-----LysGluArgHisLysLeuPheGlnGly----	285
Db	950	AGAAGCGAGGATTGTAGGCCATATTTCCGACAGCACCGCACGCGTGGGCAAAACCGCT	1003
OY	286	-TyrGlnTrhGlnGluArgGlnGluTrhGluLeuSerGlnLysIleLysLeuGlyCysG	305
Db	1010	TCTCTGACAGTCAAGCCGCGGGAAGCTTGG-----GAGGCGG	1048
OY	305	InProGluLeuSerGlnTrhSerGlnTrhLeuProProLysProPheSerCysGlyArgS	325
Db	1049	AAGCAGAGGCGGAGAGGCGGATTCGGTCCAGTCCCTGGTGGGAGGCGCCGACGCG	1108
OY	325	erGlyLysGlyHisLysArgLysSerProPheGlySerTrhGluArgLysTrhProVal	345
Db	1109	AGGCGACAGGTGAGCCACATCTACGGCC-----	1137
OY	345	ysLysLeuAlaProGluPheSerLysValLysTrhLysTrhProLysHisSerProIle	365
Db	1138	-----AAGCAGTGAACCTTCTCCGCAACCTGGGCGCGTGC	1174
OY	365	ysGlnGluProCysGlySerLeuSerGlnTrhTrpValCysLys-----	378
Db	1175	CCCAAGAGGGGCTGGGGAATACAAATCAGCGGTACGCAACCTGAGTCCCGCTTGACT	1233
OY	379	-----ArgGluLeuArgSerGlnGluTrhProG	388

Db	1235	CGCTGCAGGGACATGCAGAGGCTCAACGACACCGACGAGGCTCTGGCCCCGCGC	1294
Qy	388	1u1ysPrGArGSeSerSerValSpHrPrProPAdgLeuSerThProGlnInsgLyProS	408
Db	1295	AGGCCCCCGGATCCGCC-----CCACCACTGCTGCTGGGTATCCTTGGCCCCA	1345
Qy	408	eRThHsPrOlysgLnuLysAlaPheSer-----	417
Db	1346	CGAGCGGGGCCCAAGCTGATCCCAAGCCCCGCGGAGATCCGGCCACCGGCGCA	1405
Qy	417	-----	417
Db	1406	GCCCTGGGCCCCGCGCCAGCCAGTCACCCGACCCCGAACAACAGCGGCCAGCGAGC	1465
Qy	418	-----SerGlu1leGluSpleuProT	425
Db	1466	TGGCGCTGGCAACCTTGACACAGGGGCTTACGCCACCCACAGAGTAGACAGCTCCCC	1525
Qy	425	yrLeuSerThrThGlnuMetYrLeuCySArgTTPHISgLnPrProProSerProLeuP	445
Db	1526	ACCTG-----CGCAATTCTCCCGACAGCCAGGTCGCCGCGC	1561
Qy	445	roLeuArgLnuSerSerProLySgLnGluThrValAlaArgCysLeuMetProSerS	465
Db	1562	AGACGCGCTTCCATCTCTCTGGCGGAGAGCG-----CCCCCGAGG	1606
Qy	465	erValAlaGlyGluThrSerValLeuAlaValPro-----SerTPRArgSPHISerV	483
Db	1607	AGCCAGCGCCCGCGCGGCGCTCAACGACGATCCCGCTGGACGCTGGCAATCCCGCG	1666
Qy	483	alGlnProLeuNArgSPRoAnSPRoSerAsPleuLeuGluAsnLeuAspSerValP	503
Db	1667	CCGAGACGAGAGGACACGCGAGCCCGCTGTGTGAGAGATTCTTGAGAGAGAGCGCT	1726
Qy	503	heserLyArGHisAlaLysLeuGluLeu-----	512
Db	1727	CTTCGAGAGGGGAGCCCGCGGTGGGGGCTTCTTACAGAGTGAAGACACCTG	1786
Qy	513	--AspGluLysArgTgLySArgTTPAsP1leGlnArg1leArgLnuGlnArg1leLeuG	532
Db	1787	AGGACGAGATGGCCCAAAAGCGGCCACCTGCTGACGCGGACAGCGGAGACAGAGG	1846
Qy	532	lnArgLeuGlnLeuNArgMetYrLySLySgLy1leGlnGluSerGluProGluValT	552
Db	1847	AGGCGCGCGCGGCAAGCATGTCGAGAGGTGAGAAAGACGAGGAGGAGCGCG	1906
Qy	552	hrSerPheProGluProAspValGluSerLeuMet1leThrProPheuProV	572
Db	1907	CGAGGCTGGCCCAAG-----GAGGCCCGGCGCCAGCCCGCTGTGTCCG	1954
Qy	572	alValAlaPheGlyArgProLeuPro	580
Db	1955	CAGTCCCATGGCGACTCAGCCCT	1980
RESULT 6			
US-09-991-496-115			
: Sequence 115, Application US/09991496			
: Patent No. US20020169285A1			
GENERAL INFORMATION:			
: APPLICANT: Reed, Steven G.			
: APPLICANT: Campos-Neto, Antonio			
: APPLICANT: Webb, John R.			
: APPLICANT: Dillon, David C.			
: APPLICANT: Skelky, Yasir A.W.			
: APPLICANT: Bhatia, Ajay			
: APPLICANT: Coler, Rhea			
: APPLICANT: Probst, Peter			
: APPLICANT: Brannon, Mark			
: TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE			
: TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS			
: FILE REFERENCE: 210121.420C9			
: CURRENT APPLICATION NUMBER: US/09/991,496			

: CURRENT FILING DATE: 2001-11-20
 : NUMBER OF SEQ ID NOS: 137
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 115
 : LENGTH: 7065
 : TYPE: DNA
 : ORGANISM: Leishmania major and chagasi
 : US-09-991-496-115

Alignment Scores:

Pred. No.:	2,666-07	Length:	7065
Score:	230.50	Matches:	156
Percent Similarity:	36.60%	Conservative:	79
Best Local Similarity:	24.30%	Mismatches:	242
Query Match:	7.12%	Indels:	168
		Gaps:	30

US-10-054-935-2 (1-614) x US-09-991-496-115 (1-7065)

QY 10 AlaAlaAlaAlaProAlaGlyAsnProGluInArGLeuAspTyGluArGAlaAla 29
 Db 2616 GCTGACAGCGCGCTAGACACCGCCACGACGACGCGC-----CGAGCTGGA 2663
 QY 30 AlaLeuGlyGlyPro-----GluAspGluProGlyAlaAlaGluAlaHisPheLeuPro 47
 Db 2664 GGCACAGGTGTCAGCGCTGCGCGACGCGGAGAGCTGACACAGCG-----CCT 2714
 QY 48 ArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerGlnGlySerPro 67
 Db 2715 AGACACCGCCAC-----GACGCA 2732
 QY 68 AlaProSerProAlaGlyCysGlyGly--LysGlyArgGlyLeuLeuProAlaGly 86
 Db 2733 GCGCGCGGAGCTGAGGACGAGCGGTGGACGCGTGGCGCGCA-----CCGCGACGA 2783
 QY 87 AlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 106
 Db 2784 GCGCGCGGACGCTGCGCGCGGACGCGGAGCTGACGACGCGCTGACACCGCCAC 2843
 QY 107 ProAlaThrLysGlnAlaGly-----LleGly-----Gly 116
 Db 2844 GCAGCAGCGCGCGGAGCTGAGGACAGCTGGACAGCGGCTGCGCGGACCGGACGAGCG 2903
 QY 117 GluProAlaAlaAla-----GlyAlaGlyCysSerProArgProLysTyrgln 132
 Db 2904 GCGCCAGCAGCTGCGCGGACGCGGAGCTGACGACGCGCTGACACCGCCACGCA 2963
 QY 133 AlaValLeuProLleGlnThrGlySerLeuValAlaAlaAla-----LysGlu 148
 Db 2964 GCAGCGCGC-----CGAGCTGAGGACAGCTGGACGCGGCTGCGCGGACCGCGACGA 3017
 QY 149 ProThrProThrAlaGly-----AspLysGlyGlyAlaAlaSerProAla----- 163
 Db 3018 GCGCGCGGACGCTGCGCGCGGACGCGGAGAGCTGACAGCAGCGCTTACACCGCGCAC 3077
 QY 164 AlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyProProProLeu 183
 Db 3078 GCAGCAGCGCGCGGAGCTGAGGACAGAGTGGACGCGGCTGCGCGGACCGACGAGGC 3137
 QY 184 AlaProThrAla-----ThrAlaGlyThrLeuAlaAlaSerGlnGluArgTyrPlys 200
 Db 3138 GCGCCAGCAGCTGCGCGCGGACCGGAGAGCTGACAGCAGCGCTTACACCGCGACGCA 3197
 QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlySerGly-AlaSerSerGlnAlaAl 220
 Db 3198 GCAGCGCGCGGAGCT-----GGAGGACAGCGGTGGACAGCGGTGCGCGGACCGCGACGA 3251
 QY 220 acylsLeuLysGlnIle-----LeuLeuLeuGlnLeuAspLeuIleG1 234
 Db 3252 GCGCGCGGACGCTGCGCGCGGACCGGAGAGCTGACAGCAGCGCTTACACCGCGCAC 3311
 QY 234 uGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247

Db 3312 GCAGCAGCGCGCGGAGCTGAGGACAGAGTGGACAGCGGCTGCGCGGACCGGACGAGGC 3371
 QY 248 -----GluGluLeuLysSerGlnArgAspThrLeuVal 259
 Db 3372 GCGCCAGCAGCTGCGCGCGGACCGGAGAGCTGACAGCGCTTACACCGCGCACGCA 3431
 QY 259 argGlnGluArgMetGluArgMetGlnLeuValLysLysAspAsnGlnLysGluArg 279
 Db 3432 GCAGCGCGCGGAGCTGAGGACAGCGGTGGACAGCGGCTGCGCGGACCGGACCGGAGCGCG 3491
 QY 279 GHisLysLeu-----PheGlnGlyTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 298
 Db 3492 CCAGCAGCTGCGCGCGGACCGGAGAGCTGACAGCAGCGCTTACACCGCGCACGCA 3551
 QY 298 uLysIleLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 318
 Db 3552 GCGCGCGGAGCTGAGGACAGCT-----GCTGACGACG 3575
 QY 318 sProPheSerCysLysArgSerGlyLysGlnLysArgLysSerProPheGlySerTh 338
 Db 3576 -----GCCAGCGCTGCGCGGAGCGGACCGGACGAGGCGCGGACGAGCTGCGCG 3622
 QY 338 rGluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysTh 358
 Db 3623 CGAACCGCGAGGA-----GCTGACGACG 3646
 QY 358 rProLysHisSerProLleLysGlnGluProGlySerLysSerLeuSerGlnThrVal ----- 376
 Db 3647 GCCTAGACCGCGGACCGGACGAGCGCGGAGCTGAGGACGAGCTGCGCGGCTGCGCG 3706
 QY 377 -CysLysArgGluLeuArgSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 394
 Db 3707 CGAGCAGCGGACGAGG--CGCGCAGCAGCTGCGCGGACCGGAGAGCTGACGAGCGCG 3765
 QY 394 lAspThrProArgArgLeuSerThrProGlnLysGlyProSerThrHisProLysGlu 414
 Db 3766 CTAGACACCGCGCAGCAGCAGCGCGGAGCTGAGGACGAGCTGCGCGCGCGCG 3825
 QY 414 sAlaPheSerSerGlnLleGlnAspLeuProTyrLeu----- 426
 Db 3826 GACCGGAGGAGCGCGGCGGACGAGCTGCGCGGACCGGAGAGCTGACGAGCGCGCTA 3885
 QY 427 -----SerThrGlnMetGlyLeuCysArgTyrPheGlnProProse 442
 Db 3886 GACACCGCGCAGCAGCAGCGCGGAGCTGAGGACGAGCTGCGCGGCTGCGCGGAC 3945
 QY 442 rPro-----LeuProLeuArgGlnSerSerProLysLysGlu 455
 Db 3946 GCGGAGAGCTGACAGCAGCGCTTACAGACCGCGGACGAGCGGCTGCGCGGAGCTG 4004
 QY 455 uThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAla 475
 Db 4005 ACGGCTGGCGACG-----CT 4019
 QY 475 lProSerThrArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeu 495
 Db 4020 GCGCGGAGCGGAGGAGCGCGGACGAGCTGCGG--GCCAGCGCGGAGAGCTGCA 4076
 QY 495 uGlnAsnLeuAspSerValPheSerLysArgHisAlaLysLeuGlnLeu----- 512
 Db 4077 GCAGCGCTTACAGACCGCG-----ACGCGACGCGCGCGGAGCTGAGGACGAGGTCGC 4130
 QY 513 -----AspLysArgArgLysArgTyrAspIleGlnArgIleArg 526
 Db 4131 ACGGCTGGCGCGGAGCGGACGAGAGGCGCGGACGAG-----CTGCGCGGAA 4178
 QY 526 gGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetTyrLysLysLysGlyIleGlnG1 546
 Db 4179 CGCGGAGAGCTGACAGCGCTTACAGACCGCGGACGAGCAGCGCG-----GCCGA 4229
 QY 546 uSerGluProGlnValThrSerPhePheProGluProAspAspValGluSerMetI1 566
 Db 4230 GCTGAGGACACAGTGTGACGCTGCGCGGAGAGCGCGGAGAGCTGACGAGCGCTTGA 4289


```

QY 455 uThValAlaArgCysLeuMetProSerSerValAlaGlyLThrSerValLeuAlaVala 475
Db 4005 ACCGGGTGCACGG-----CT 4019
QY 475 lProSerTrpArgSpHisSerValGluProLeuArgAspProAsnProSerAspLeu 495
Db 4020 GGGCCGGACGCGACAGAGGGCGCCAGCAGCTGGC---CGGAACCCGAGAGGCTGCA 4076
QY 495 uGluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGluLeu----- 512
Db 4077 GCACGCGCTAGACACCGCC-----ACGACAGACGGCGGAGCTGAGAGGACGGGTGGC 4130
QY 513 -----AspGluLysArgArgLysArgTrpAspIleGlnArgIleAr 526
Db 4131 ACGGCTGCGCGGACCGCGCGACGAGCGCGCCAGCAG-----CTGGCCGGCAA 4178
QY 526 gGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetLysLysLysGlyIleGlnI 546
Db 4179 CGCCGAGAGAGCTGACAGACGCGCTAGACACCGCCAGCAGCAGCGC-----GCCGA 4229
QY 546 uSerGluProGluValThrSerPheProGluProAspAspValGluSerLeuMetI 566
Db 4230 GCTGGAGGACACAGGTGCACGGCTGGCGCGGACGAGCGAGAGCTGCACGCGCTAGA 4289
QY 566 eThr 567
Db 4290 CACC 4293

```

RESULT 8

US-10-037-270-713

Sequence 713, Application US/10037270
 Publication No. US20030104529A1

GENERAL INFORMATION:

```

: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyan
: APPLICANT: Chen, Rui-hong
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wehrman, Tom
: APPLICANT: Xue, Aldong J.
: APPLICANT: Yang, Yonghong
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Ma, Yungqing
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhivael
: APPLICANT: Tillinghast, John
: APPLICANT: Dimauc, Radoje T.
: TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
: FILE REFERENCE: 784CIP2B
: CURRENT APPLICATION NUMBER: US/10/037,270
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ. ID NOS: 1104
: SOFTWARE: pt_genes Version 1.0
: SEQ ID NO 713
: LENGTH: 3169
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (116)..(2146)
: US-10-037-270-713

```

Alignment Scores:

Pred. No.: 1.61e-07 Length: 3169
 Score: 228.50 Matches: 151

```

Percent Similarity: 36.25% Conservative: 77
Best Local Similarity: 24.01% Mismatches: 199
Query Match: 7.06% Indels: 204
Db: 9 Gaps: 32
US-10-054-935-2 (1-614) x US-10-037-270-713 (1-3169)
QY 13 AlaProAlaGlyLysAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
Db 47 AGTCCGCTCCCGGCGCTCG----- 67
QY 33 GlyProGluAspGluProGlyAlaAlaGlnAlaHisPheLeuProAlaGHisArgLysLeu 52
Db 68 TCACACATGATGATCCAGCCCGCGGAGCGCCGCGCGCCGCGCATGCCCTCGCC 127
QY 53 LysGluProGlyProProLeuAlaSerSerGlnGlyLysSerProAlaProSerProAla 72
Db 128 AATCAAGAGGCG-----TCAAGGCGCGCCACAGCGCGCGCGAGCCCTCG 172
QY 73 GlyCysGly-----GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro 89
Db 173 GAGAAAGGTGCCACCCGTCGCGCGCGCGCTGCTCGCGCGAGCAGCAGCGCC 232
QY 90 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 109
Db 233 GCGTCCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 283
QY 110 LysGlnAlaGlyLysGlyGlyGluProAlaAlaAlaAlaGlyCysSerProArgPro 129
Db 284 GCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
QY 130 LysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaAlaLysGluPro 149
Db 344 -----TCGCCCAACAA- 355
QY 150 ThrProTrpAlaGlyAspLysGly-----GlyAlaAlaSerProAlaAlaThrAlaSer 167
Db 356 -----AGGGCTCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 406
QY 168 AspProAlaGly-----ProProLeuProLeuProGly 179
Db 407 CACCCGTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 466
QY 180 ProProProLeuAlaPro-ThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArgTr 199
Db 467 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
QY 199 PylSerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 214
Db 524 GCGCAGCGCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
QY 215 -----AlaSerSerGlnAlaAlaLysLeu 223
Db 583 CTCCTCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
QY 223 sGlnIle-----LeuLeuLeuGlnLeuAspLeuIle----- 233
Db 643 CAGGCTCGGAGGCGCTCAACTTCTCTCACTGCGCTGCTGCGCGCGCGCGCGCTT 702
QY 233 ----- 233
Db 703 CTCGGGCTGGTGCCTCCACAGCTCTGAGAGAGTCCACAGGTCCGGCGGACCA 762
QY 234 -----GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
Db 763 GGAATTCTCCCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 822
QY 247 eGluGluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArg 267
Db 823 GCACTCTTGCAGGCAATTTGGAATTTGAGTCACTTGTAGAAAGTCCCAACATA 882
QY 267 gMetGlnLeuValLysLys-----AspAsnGluLysGluArg---HisLys 281
Db 267 ----- 281

```

Db 883 ACAAGACCTCAGAGAAAGCTGTGAAAGAGGGAGAGTGAAGCTACGCCGATCAGCGA 942
Qy 281 sleuPhengInglYtRgluThrGlu----- 289
Db 943 AGTGTCTGAGAACTCCAGATGAGATTCTCAAGACCTCTCGGATGGGATTCATGTGCT 1002
Qy 290 -----gluArgIngluThrGluLeuSerGluIleuSerIleuLeuGluGlu 305
Db 1003 GAAGAGACCGCGGGAGCGGAGCTTCAGCTCCCTGGAGAACACGGTGGAG---GAGCGGCT 1059
Qy 305 nProGluLeuSerGluThr--SerGlnThrLeuProProGluProPhSerGluGlu 325
Db 1060 GACGAGCTCAACCAATCCATCAACGACGACATCGCA-----TCTTCACGAGAGT 1110
Qy 325 erGluYsGluHIslyuArgLysSer-----ProPhe----- 335
Db 1111 CCAGAGAGAGCCCAAGAGAGATCATGACATGAGGCAAGAGTTGCTCCCTGGAGAGA 1170
Qy 336 -----GlySerThrGluArgLysThrProValLysLysLeu----- 347
Db 1171 ATCTGAGGGGAGAACAGAGATTTGAAGCCTTAAGAGACTGTGAAGAGATACAGAC 1230
Qy 348 -----AlaProGluPheSerLysValLys----- 355
Db 1231 CTCAGCCAACTCCAGAGAGTGGACATGGAGCGCCCTGAGAGATACCTTCAGACTATGA 1290
Qy 356 -----ThrLysThrProLysHisSer----- 362
Db 1291 GTCTGACATCTACACCGAGGCTCCGAGCTGCTGAGCTCAAGCAGGAGCAGCGCTTT 1350
Qy 363 -----ProIleuYsGlu---GluProCysGlySerLeuSerGluThrValCysL 378
Db 1351 CAAGAGAGCGGCGGACACGAGCGGCTGCGCCCTGAG--GCCCTCAGCGAGAACTTCACA 1409
Qy 378 ysaArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProp 398
Db 1410 GGCTGTGAGAGCTCCGCTCCCGCTCCGAGGAGATCCGAGACTGGAGAGAGAGCTCC 1469
Qy 398 roaArgLeuSerThrProGlnLysGlyProSer-----ThrHisProLysG 413
Db 1470 GCCAGCTGAAGTCCGATTCACAGCGGCCGAAAGAGAGAGCGCTTCAGACACTCGGAG 1529
Qy 413 lu---LysAlaPhe--SerSerGluIleGluAsp---LeuProTYrLeuSerThrGlu 430
Db 1530 CCTTGAAGCAGCTCCAGAAAGAGTCAAGGAGCTCCAGGCTCCAGCAGCGGAGG 1589
Qy 431 MetIyLeu-----CysArgTTP-----HisGlnPro--ProProSerProLeuPr 445
Db 1590 ATGGGCTGCTCTCCATGAGGTGCTCTGCGCGCCAGAGAGAGCTTGAGATCCCTCC 1649
Qy 445 oleuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSe 465
Db 1650 TGTCTCAGAG-----CCAGAGACGACGAGCGGCTGCGCC-----CCTGCAGG 1694
Qy 465 rValAlaGlyLysLysThrValLeu---AlaValProSerTTP---ArgAspHisSerVa 483
Db 1695 GCGCGCTGGAAGGCGCTGGGTCTTCAGAGGAGACACAGATGGCTGCGCCAGCAGGGA 1754
Qy 483 lGluProLeuArgAspPro 489
Db 1755 GGAGCTTGGGAGACCA 1773

RESULT 9

US-10-171-311-51
; Sequence 51, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kametkar, ShubhangI
; APPLICANT: Glatz, Karen

; APPLICANT: Gannayapur, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIORITY FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 15952
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-171-311-51

Alignment Scores:
Pred. No.: 1.19e-06 Length: 15952
Score: 226.50 Matches: 172
Percent Similarity: 33.10% Conservative: 64
Best Local Similarity: 24.12% Mismatches: 264
Query Match: 7.00% Indels: 213
Gaps: 33
DB:

US-10-054-935-2 (1-614) x US-10-171-311-51 (1-15952)

Qy 9 LysAlaAlaAlaAlaAlaAlaAlaGlyLysnProGluIngluArgLeuAspTYrGluArgAla 28
Db 10408 CGTGAAGTGTGCTGCGCCCTGCTGCTGCGGACCGGACACCGGCTCTACATGTGACACT 10467
Qy 29 AlaAlaLeuGly-----GlyProGluAspGluProGlyAlaAlaAlaAlaHisPhe 45
Db 10468 GGCAGCGCGGGGCTCGCGCTGACGACCTGAGCAGAGAGAGTGGCTGCGCTCGCGGA 10527
Qy 46 LeuProArgHisArgLysLeu---LysGluProGluPro----- 57
Db 10528 CGCCGCTGACGCGGAGCTGCGGCGCCCTCAGGCGCAGAGCGTCTCCGTGGAGCT 10587
Qy 58 ProLeuAlaSerSerGlnGlyGlySerProAlaProSerPro----- 71
Db 10588 CCTCTTCAACCGCAGGTGTCCGAGAGACCGGCGGAGACCTGCTGACAGATACCGGCG 10647
Qy 72 -----AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro 89
Db 10648 GGGCAGCGTGAACGTGGAGAGTGGGCGCACCTCACCTGCTGCTGGCCAGGCCCA 10707
Qy 90 GlyGlnGlnGluGluSerTTPGlyGlySerValProLeuProCysProPro-----Pro 107
Db 10708 GCGCCAGCGCGGCGGCGGAGCGGCGGAGCGGCGGCGGAGCGGCGGCGGAGCGCT 10767
Qy 108 AlaThr---LysGlnAlaGlyTLeGlyGlyGluProAlaAlaAlaAlaGlyLysSer 126
Db 10768 GGTGCGGCGCACCATGAGGTCAAGGTGGCGGCTCCGGGGCG----- 10812
Qy 127 ProArgProLysTYrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaAla 146
Db 10813 -----CGCGGCGCGGTGGAGAGTCT 10836
Qy 147 LysGluProThrProTTPAlaGlyLysPlyGlyGlyAlaAlaAlaSerProAlaAlaThrAla 166
Db 10837 GCGCTCGGCTACGTGACGCGGCGCGCCCGGAGAGAGCTGCTGGCGGAGTTGG----- 10890
Qy 167 SerAspProAlaGlyProProProProLeuProLeuProGlyProPro----- 181
Db 10891 -----CTCGGGAGACCTTGAGCTTCCCGCGGTGACCGCGGCTGACCGCGCATCAT 10941
Qy 182 -----ProLeuAlaProThrAlaThrAlaGlyTYrThrLeuAla----- 193

Db	10942	CGAGGAGCGCCGAGGAGGCCCCCGGGGCGCGCGCGCGCGAGCTCCAGAGAGCGCTGGGCGCGCC	11001
OY	194	-----AlasergluglyArgTrpLysSerMetArgLysSerProLeuagllyg	209
Db	11002	CGGGAGCCAGAGGCCACCGGGCGAGGGAGCGAGCTCGGGGCGCTCCACAGCAGAGG	11061
OY	210	gllyllygly-----SerGly-AlaserSerGlnAlaAlaCysLeuylsGlnI	225
Db	11062	CCAGGGGGAGGGCGGAGACCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGA	11121
OY	225	eLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnLeuGlnAlaLysGlu	245
Db	11122	GCAGACCCCTGGCTATGCCACATGAGAGGTGCAGCGGGGCGAGTTCACAGGGCGGCGGT	11181
OY	245	sgLIllleGluLeu-----LysSerGluArgAspThrleu	258
Db	11182	CTCCGTGGGAGCTCTTCTTCTCTGTAACCTGAGCGAGGCGCGCGAGAGAGCTCT	11241
OY	258	uAlaArg-----	260
Db	11242	GGCCACAGCGCGCGCGCGCGCGCTGGGCTGCCGACCTGTCGCGCTCACCGGGGT	11301
OY	261	-llleGluArgMetLuarArgMetGlnLeuVal-----LysLysAs	274
Db	11302	CATCGAGGAGCGGAGGAGCGGCTCAGCAAGGTCTTCGCGGCTGAGGGCGAGGT	11361
OY	274	pAsnGluLysGluArgHis-----LysLeuP	283
Db	11362	GTCGCGCTCGAGGTGCACACGTCCGGAGTCTGGGCGCGAGACCTCGGGAGCTGGC	11421
OY	283	eGlnGlyTYrGluThrGlnGluArgGlnThrGluLeu--SerGluLysIleLysLe	302
Db	11422	CCAGGGGCACTAGACGCTGCAG--GAGGTGACGAGGATGTGACTCGTCAACGCTACT	11478
OY	302	uGlu-----CysGlnProGluLeuSerGlnThrSerG	313
Db	11479	GGAGGGCACCAAGCTGCATCGGGGCGTCTGTGTGCCCGCCAGAGCACGCGGCGCGCA	11538
OY	313	lnThrLeu-----ProLysProP	320
Db	11539	GAGAAATGATGAGCTATTCACGAGCATGATGGAAGGCGGTCTCGGCGCGGACAGGCTCT	11598
OY	320	heserCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThr	338
Db	11599	GGTCTGCTGAGG-----CGAGGGCGCCACCGGCTTCATGCACCC	11643
OY	338	-----	338
Db	11644	CGTCCGCACTGAGGCTGTGCGTGGAGGAGGCGCGTGGCGCGGCGGTGGCGCGCA	11703
OY	339	--GluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysT	358
Db	11704	GATCCAGGAAGAGGTGCTGTGCGGCGGAGCGCGCG--TCACCGGCTA	11748
OY	358	hrProLysHisSerProIleLysGluLeuProCysGlySerLeuSerGlnThrValCysL	378
Db	11749	CACCGACCCCTACACCGGCGAGATCTCC-----TCTTCAGGCCATGGA	11796
OY	378	ysArgLysLeuArgSerGlnGluThr-----ProGluLysProArgS	392
Db	11797	GAAAGACCTGATCGTCCGGAGAGCAGGACATCCCGTGTGAGAGCCCAATGCACCGG	11856
OY	392	eIserValaPrrProProArgLysSerThrProGlnLysGlyProSerThrInHisProL	412
Db	11857	CGGGGTATCGACCCG-----TGCAACACCAACCGGTGCGGTGAGCTGGCTG	11907
OY	412	ysGluLysAlaPheSerSerGluIleGluAspLeuProTyLysSerThrInGluMet	432
Db	11908	CCGGCGGGCT-----ACTTCAGCAGAGAGAT	11934
OY	432	ylLeuCysArgTrpHisGlnProProPioSerProLeuArgLysSerProL	452

[illegible]

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3117 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 US-09-752-639-6

Alignment Scores:
 Pred. No.: 3,59e-07 Length: 3117
 Score: 223.50 Matches: 185
 Percent Similarity: 31.67% Conservative: 76
 Best Local Similarity: 22.45% Mismatches: 210
 Query Match: 6,908 Indels: 353
 DB: 10 Gaps: 42

US-10-054-935-2 (1-614) x US-09-752-639-6 (1-3117)

QY 12 AlaAlaProAlaGlyGly-----AsnProGluGlnArgLeuAspTyrGluArgAla---- 28
 DB 293 GCTGTCCAGCAGGTGGCTCAGTAAAGTGGCCCAACTGTGTGATGGCTCCAGGCGGG 352
 QY 29 -----AlaAlaLeuGlyGlyPro 34
 DB 353 CCCGAGCGTGGAGAGGTGGGGTGCATGACAGCAGCTGGCAGCAGCAGCAGCAGCAGC 412
 QY 35 GluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgLysLeuLys--- 53
 DB 413 GCTTCACCCCATTCACATGAA-----CTGCCACAGTCTGCCCTCTACAGTGC 463
 QY 54 -----GluProGly-----ProProLeuAla 60
 DB 464 AACCAAGGGAGCCCGCATCTCGAGTGGAGTCCCGCACTTACTATACCACTCGTAGGC 523
 QY 61 SerSerGlnGly-----GlySerProAlaProSerPro----- 71
 DB 524 ACTGAAGCGGAGAAAGGGGGGGCCACAGCTGGACCGCATGTGGCAGCAATGATGCC 583
 QY 72 -----AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro 89
 DB 584 ACAGAGGTGCAGCTGAGGTAGG-----GCGGCCCA 616
 QY 90 GlyGlnGlnGlu-----GluSerTyrGlyGlySerValProLeuProGlySerProProAla 108
 DB 617 GGCACCCCTGAAATTTCTTCCACGACGCCAAGAAACCCCAACCAAGTCACCTGCCCTGCA 676
 QY 109 ThrLysGlnAlaGlyLys-----GlyGlyGluProAlaAlaAlaGlyAlaGly---Cys 125
 DB 677 ACCCTTCAGCTGGCATTTGGCGCACAGGTAAACGGCAGGTCTTCGGCAGGGGCCAC 736
 QY 126 SerProArgPro-----LysTyrGlnAlaValLeuProIleGlnThr 139
 DB 737 GCCCCCAAAACCGGTGGCTGCTTCCCTCCACAGAACAGCAGCAGCAGCAGCAGCAACA 796
 QY 140 GlySerLeuValAlaAlaAlaLysGluProThrPro----- 151
 DB 797 GCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 856
 QY 152 -----TrpAla 153
 DB 857 TTTCATGCCACAGCAACCTTCGACGACCAACCCAGAGCTTTGGCTGACGACGAGTGGGCC 916
 QY 153 ----- 153
 DB 917 ACTGGAGAGTCCACCTGGCTACACAGCAGATGACACCTTACCTCCGCCCAACCC 976
 QY 154 -----GlyAspLysGlyGlyAlaAlaSerProAlaAlaThr 165

DB 977 AGATATGAACCCAGAACTGCGCAAGGCCCTTCTCAGAGA---CTACAGCCCGCAGCCAGC 1033
 QY 166 AlaSerAspProAlaGlyProProProLeuProGluProGlyProProLeu----- 183
 DB 1034 GCTACG-----TAGGTCCAGATCCCTTCCCGCGCGCTCCCGGCTCTCTAAGGA 1087
 QY 183 ----- 183
 DB 1088 GGTATCCTGCTCCAGAGCGCTGTGATGGGGCTGGCACCAGCCTGGCAGAGGCCAC 1147
 QY 183 ----- 183
 DB 1148 TGGCAACCTGTTCTATACATCAGTGGCCCTGACAGCAGCCGACCTGCTCTGGGGCA 1207
 QY 184 AlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGly----- 197
 DB 1208 GCCCATTCCTGACCTCTGGATTCCTCCGCTGGAGCTGAGGAGTGCAGCTACTGCTGA 1267
 QY 198 -----ArgTyrLysSerMetArgLysSerProLeuGlyGlyGly 211
 DB 1268 TGGGAGAGACTACACCCCATATGGCGGAGCGAGAGGCTCTCCATGGGACAGAGGA 1327
 QY 212 GlySerGlyAlaSer-SerGlnAlaAlaCysLeuLysGlnIleLeuLeuGlnLeuAs 231
 DB 1328 GGGCATGAGGGCAGTGTGACAGAGGGAGCTGTGGCAG-----GTGCTACGGGGCG 1378
 QY 231 PheIleGluGlnGlnGlnGlnGlnGlnAlaLysGluLysGluIleGluGluLeuLys 251
 DB 1379 AGTATTCAGAGCAGCAGGAGGAGGAGCCCGGCA----- 1412
 QY 251 SerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgMetGlnLeuVa 271
 DB 1413 -TCCAGAGAGCCCAATTTGCTGACCC-----CTGCCACAGAGCTGTGGAGCTGCG 1462
 QY 271 LysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTyrGluThrGluGlu 291
 DB 1463 CTCACCTCAGATCAAGAT-----GCGAGTGGCTCTGAAGAA 1504
 QY 291 gGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGlu 311
 DB 1505 GCGCAAAAGTGTATTTGGCTCCTCACTACCAAG-----TGTGGGTGAGTTTCTGAGCC 1558
 QY 311 rSerGlnThrLeu-ProProLys-----ProPheSer- 321
 DB 1559 TTCC-----TTACCCACCAAGCAGCAGCAGAGACAGTGGATGTACCTCCATCAT 1612
 QY 322 -----CysGly----- 323
 DB 1613 CCACAGTCTGTGCTGTGCAACTGTGACCCCACTGAGGACAGCCAGCTGGAGCTCT 1672
 QY 324 -----ArgSerGlyLys-----GlyHisLysArgLys 333
 DB 1673 TGATGAGAGAGGGAGGAGTCTTGAACAGAACCTCTGAGACACAGCATCATCATCGT 1732
 QY 333 erProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSer 353
 DB 1733 CACCCGAGCGCGGTCCACCC-----GAATCCCGGGCAGCAGTCTC-----A 1774
 QY 353 yValLysThrLysThrProLysHisSerProIleLysGluGluProCysGlySerLys 373
 DB 1775 AGCTCAGGCGGAGGACATGATGTCAGAGTTGGAGGGGAGACCTTCCGTGGGAAACCAA 1834
 QY 373 er-----GluThrValCysLysArgGluLeuArgSerGlnGlu 386
 DB 1835 GCAAGCGGCCAGGCGCAGCCCATCATCATCATCATCATCATCATCATCATCATCATCAT 1894
 QY 386 hrProGluLysProArgSer-SerValAspThrProProArgLeuSerThrProGluLys 405
 DB 1895 TCCCGTCTACTCCAAATCATCAACCATCATCATCATCATCATCATCATCATCATCAT 1945
 QY 406 GlyProSer-ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuPro 425
 DB 1946 GCGCTAGCTGACCAACCCCTGTGAGCGGAGCTTT-----GAGCTACCTCC 1990

```

QY 425 rleuserthrthrlumetyrleucysargtrphsleuinprouproserproleupr 445
Db 1991 C-----TACAGCGCGCCCGCCATCCTCAGCC 2017
QY 445 oleuargluser----- 450
Db 2018 TGTGGGAAAGGCTGTGCTCTACTTCAATGCCATCATATCAACAGCACCACCTGCTGC 2077
QY 451 -----Prolysylsleuigluthrvalalalargcysleumetproser----- 464
Db 2078 CCCTCTCCCATCAAGCGCTAGAGTGCATCGACGCTGCTCCGAGCTAACAGTGTGTA 2137
QY 465 -----SerValAlaArgLysThrSerValLeu----- 474
Db 2138 AGTAAACCCCGCTCTCTCTGTGTGTATGGGGAGGCCACCCCATGAGCATCGAGCCAG 2197
QY 475 -----ValProSerTrpArgAspHisSerValGl 484
Db 2198 GATCAAGTGGGCTCCGGTTCAGCAGCAAGAAATCCCTGTATGAGGACCGCTGCCCTG-- 2255
QY 484 uProleuArgspProAsnProSerAspLeu-----GluAsnLeuAspAs 500
Db 2256 -GCAGCTGCAGATCCCAAGAGCTGACTGTGTGGAGCCATGAGGAGCAAGAGAG 2314
QY 500 pSerValPheSerLysArgHisAlaLysLeuAluLeuAspLysArgTrp 520
Db 2315 CAGC-----CGGAGAAAGCAGAGGCAAGTGA 2341
QY 520 PASP----- 521
Db 2342 AGACCTGCTGCACAGCCGCTGCTCCAGCATTTCTGCTGCTGCGACCAACAGAGACT 2401
QY 522 ----lIeGlArgLleArgLgluArg-----lIeLeuAlnArgLeuAlnArgMe 538
Db 2402 GGCCCTGCACGTGTGCAGCATCCAGAGAGAGACATCTGTGAACCGCTGAATTAAGCTGCT 2461
QY 538 tTyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 558
Db 2462 GCTGAAGAAAGCCCTGGCGGCCCAACATCCCTGCAACTTATCTACACAGGCTC 2521
QY 558 oASP 559
Db 2522 TGAC 2525

```

RESULT 12

```

US-09-984-198-6
; Sequence 6, Application US/09984198
; Patent No. US2002010679A1
; GENERAL INFORMATION:
; APPLICANT: Gatanaga, T.
; TITLE OF INVENTION: Factors Altering Tumor Necrosis
; TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods
; NUMBER OF INVENTIONS: of Use Thereof
; NUMBER OF SEQUENCES: 154
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/984,198
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: PCT/US99/10793
FILING DATE:
APPLICATION NUMBER: 09/081,385
FILING DATE:
APPLICATION NUMBER: 08/964,747
FILING DATE: 05-NOV-1997
APPLICATION NUMBER: 60/030,761
FILING DATE: 06-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wu, Frank
REGISTRATION NUMBER: 41,386
REFERENCE/DOCKET NUMBER: 22000-20577, 21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3117 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-984-198-6

```

```

Alignment Scores:
Pred. No.: 3,59e-07 Length: 3117
Score: 223.50 Matches: 185
Percent Similarity: 31.67% Conservative: 76
Best Local Similarity: 22.45% Mismatches: 210
Query Match: 6,90% Indels: 353
DB: 10 Gaps: 42

```

US-10-054-935-2 (1-614) x US-09-984-198-6 (1-3117)

```

QY 12 AlaAlaProAlaGlyGly-----AsnProGluAlnArgLysAspTrpGluArgAla--- 28
Db 293 GCTGCCAGCAGAGTGGGCTCAGTAAAGTGGCCCAACTGTGTAGTGGCTCCAGGGCGGG 352
QY 29 -----AlaAlaLeuGlyGlyPro 34
Db 353 CCCGAGCGTGGAGAGGTGGGGGTGTCAGTACAGACAGCTGGAGCAGCAGCGCAGGCA 412
QY 35 GluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgLysLeuS--- 53
Db 413 GCCTCCACCCCATTCACATGAA-----CTGCCACAGTCTGTCTCTTACAGTGC 463
QY 54 -----GluProGly-----ProProLeuAla 60
Db 464 AACCAAGGAGAGCCCGCATCTGGAGTGGAGTCCCGACTTACTATACACACCTGAGGC 523
QY 61 SerSerGlyGly-----GlySerProAlaProSerPro----- 71
Db 524 ACTGAAGCGGAGAAAGCGGGGCCACAGCTGACCGCTATGTGCGACCAATGATGCC 583
QY 72 -----AlaGlyCysGlyGlyArgLysGlyLeuLeuProAlaGlyAlaAlaPro 89
Db 584 ACAGAGGTGACGCTGGAGTAG-----GGGCGCCA 616
QY 90 GlyGlnGlnGlu-----GluSerTrpGlyGlySerValProLeuProCysProProProAla 108
Db 617 GGCAACCCCTGAATCTTCTCCACGACGACCAAGAAACCCCAACCAAGTACAGCCCTGCA 676
QY 109 ThrLysGlnAlaGlyIle-----GlyGlyLeuProAlaAlaAlaGlyAlaGly---Cys 125
Db 677 ACCCTTCAGCTGGCATTCGGCCACCAAGTGAACCGGAGGTCTTCCGGGGGCCACC 736
QY 126 SerProArgPro-----LysTrpGlnAlaValLeuProIleGlnThr 139
Db 737 GCCCCCAACCCGGGTGGCTGCTTCTCCACAGAGAGCAGCAGCAGCAGCAACCA 796
QY 140 GlySerLeuValAlaAlaAlaLysGluProThrPro----- 151

```


Db	797	GCAGAGCAGCAGCAGCAGCAGCAGCGCCCTAACCCAGATGCCCGCTTTTGAAACTCTGA	856
OY	152	-----	-trpala 153
Db	857	TTCCATGCCACAGCAACCCCTGCAGCAACCCCGAGACTTTGGCTCAGACGACGCTGGGCC	916
OY	153	-----	----- 153
Db	917	ACTGGGACAGTCCACCTGGCTGACACCAAGATGGCACCCTTAACCCCTTCCCCCAACC	976
OY	154	-----	-glyaspysglglyalalaserproalalathr 165
Db	977	AGATATGAACCCAGAACTGGCGAAGGCCCTTCTGAGAGA-----CTGAGCCCGCAGCCAGC	1033
OY	166	AlaserapProalaglyProProProleuProleuProglyProProProleu-----	183
Db	1034	GCTACC-----TCAGGTCCAGATCCCTTCCCGCGCCTCCGCGCCCTCTCTPAAAGA	1087
OY	183	-----	----- 183
Db	1088	GGGATACTGCTCCACAGCCCGCTGGATGGGCTGGACCCAGCTGGGAGAGGCCAC	1147
OY	183	-----	----- 183
Db	1148	TGGCAACCTGTTCTACATCACTGGCCCTGCAGAGCGCCCACTGGCTCCTCTGGGCA	1207
OY	184	AlAprothralathrAlaglyThreualaAlaserglyly-----	197
Db	1208	GCCCCATCCTGAAGCTCTGGATTCCCTCTGAGCTGAGGAGTCGACACTACCTCGA	1267
OY	198	-----	-ArgTPrpyssermetarglysserProleuglyglyly 211
Db	1268	TGGGAGAGACTAGCACCCAAATGGCCGGAGCGAGAGGCTCTGCCATGGCGACGAGGA	1327
OY	212	glyserglyalaser-SerginalalalacysleuylslnleuleuleuLeuLeuLeuAs	231
Db	1328	GGGCGATGAGGACAGTGAACACAGGGAGACTGTGGGAG-----GTGCTACGGGGGG	1378
OY	231	pLeuileuglInglInglInglInleuGlAlalalysglulysglulileuglulLeuLy	251
Db	1379	AGTATCCAGAGCACGCGACGAGCGCGCGGGCA-----	1412
OY	251	sSergluarapsThreuleuLeuAlaArgllleuargmetgluarargmetglInleuVa	271
Db	1413	-TCCAGAGGCCCAATTGCTGACC-----CTGGCCAGAAAGCTGTGGACTTGGC	1467
OY	271	llylelylsapasnlglysluarargHslslyleuPheglnglyTyrgluthrgluar	291
Db	1463	CTCACTCGAGATGCAAAGAT-----	-ggCACTGGTTCGAAAGAA 1504
OY	291	gglugluthrgluenserglyllysllelylsleuGlncysglInprogluleuSergluth	311
Db	1505	GCGGAAAGTATATATGGCTCACTACCAAG-----TGTGGGNGAGATTTTCTGAGCC	1558
OY	311	rSerglInThreu-ProProlys-----	-ProPrieser 321
Db	1559	TTCC-----TTAGCCACCAAGCAGACGAGAAAGACAGTGAGATGTACCCCTCATAT	1612
OY	322	-----	-cysgly----- 323
Db	1613	CCCAGTGTCTGTGCTGTGCCAACTGTGGACCACTAGGACGCCAGCTGGAGGCT	1672
OY	324	-----	-ArgSerglylys-----
Db	1673	TGATAGAGACGCGGAAGGCTTTGAACAGAACCTGTGTCAGCAACAGCATCATGATCATGT	1733
OY	333	erProPheglySerThrgluarglyThrProvalylslyleuAlaProclunPheSerL	353
Db	1733	CACCCGAGGGGCTCACACC-----GATATCCCGGAGCAATGCTC-----	-A 1774
OY	353	ysValylsThlysthrProlyshsSerProIlelysglIngluProcysgllyserLeus	373
Db	1775	AGCTCAGCGGAGCATGATGTGCATGTTGGAGGGGAGGACCTTCCGTCGGGAAACAA	1834

```

QY 373 -----GluThrValCysIysArgGluLeuArgSerGlnGlu 386
Db 1835 GCAGCGGCCCAAGGCCCGGAGCCCTTCATTCATCCCAACCAAGAGGGGACCTTTCATCGCCCC 1894
QY 386 hrProGluLysProArgSer-SerValAspThrProProArgGluSerThrProGlnLys 405
Db 1895 TCCGCGTACACCAACATCACCACCATACCAAGAGCACCTGGCGTCTC-----CGT 1945
QY 406 GlyProSer-ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTy 425
Db 1946 GCGCGTACTACCAACACCCCTCTGAGCGGAGCTTT-----GAGTACACTCC 1990
QY 425 rLeuSerThrThrGluMetTyrlLeuCysArgTrpHisGlnProProProSerProLeuArg 445
Db 1991 C-----TACACGCCGCCGCCCATCTCAGCCC 2017
QY 445 OLeuArgGluSerSer----- 450
Db 2018 TGTGGGGAAAGGCTCGTGCCTCTACTTCATATGCCATCATATACCAACAGACCATCTCTGC 2077
QY 451 -----ProLysLysGluGluThrValAlaArgCysLeuMetProSer----- 464
Db 2078 CCGTCCTCCCATCAACGAGCCTTAAGATGCCATCAGCAGCGTGTCTCGGACTACAGTGTGA 2137
QY 465 -----SerValAlaGlyGluThrSerValLeuAla----- 474
Db 2138 AGTAACCCCGGCTCTCTCTGTGTGTATGGGGAGGAGCCACCCCACTGACATCGAGCCACG 2197
QY 475 -----ValProSerTrpArgAspHisSerValGlu 484
Db 2198 GATCAACGTGGGCTCCCGGTTCCAGAGAGAAATCCCTTGATGAGGGAGCGTGCCCTG-- 2255
QY 484 uProLeuArgAspProAsnProSerAspLeuLeu-----GluAsnLeuAspAs 500
Db 2256 -GCAAGCTGCAATCCCAACAAAGCTGACTGTGTGTGGCAGCAGCATGAGGAGGACTTAAGAG 2314
QY 500 pSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLysArgLysArgTr 520
Db 2315 CAGC-----CGGAGAGAGCAGAGGCAAGTGA 2341
QY 520 pAsp----- 521
Db 2342 AGACCTGCTGCAGACCGGCTGCTCCAGCATTTTCCCTGTGCTGCACCAACGAGAGCT 2401
QY 522 ----IleGlnArgIleArgGluGlnArg-----IleGlnArgIleGlnLeuArgMe 538
Db 2402 GCGCCCTGACTGTCTGCAGATCCAGATCCAGGAGACATCCTGGAAACGCTGAATTAAGCTCT 2461
QY 538 tTyrlLysLysLysGlyIleGlnGluSerGluProGluValThrSerPheSerProGluTr 558
Db 2462 GCTGAAGAAAGCCCTGGGGGCCCAACACCATCGCTGGCAACTTATCATCAACAGGCTC 2521
QY 558 oAsp 559
Db 2522 TGAC 2525

RESULT 13
; US-09-563-728A-31
; Sequence 31, Application US/09563728A
; Publication No. US20030078216A1
; GENERAL INFORMATION:
; APPLICANT: Macleod, Alan R
; APPLICANT: Li, Zoumei
; APPLICANT: Besterman, Jeffrey M
; TITLE OF INVENTION: Inhibition of Histone Deacetylase
; FILE REFERENCE: 106101.229
; CURRENT APPLICATION NUMBER: US/09/563.728A
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/132,287
; PRIOR FILING DATE: 1999-05-03
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.1

```

RESULT 13
 US-09-563-728A-31
 ; Sequence 31, Application US/09563728A
 ; Publication No. US20030078216A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Macleod, Alan R
 ; APPLICANT: Li, Zoumei
 ; APPLICANT: Besterman, Jeffrey M
 ; TITLE OF INVENTION: Inhibition of Histone Deacetylase
 ; FILE REFERENCE: 106101.229
 ; CURRENT APPLICATION NUMBER: US/09/563,728A
 ; CURRENT FILING DATE: 2000-05-03
 ; PRIOR APPLICATION NUMBER: 60/132,287
 ; PRIOR FILING DATE: 1999-05-03
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.1

90

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 867 Seconds
(without alignments)
11356.098 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagctcctgcgccccgcgcac.....aaaaaaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1478	33.8	1509	AA224879	Human secreted pro
2	837	19.1	1529	AA18193	Lung cancer associ
3	785.4	18.0	2887	AA18193	Human ORF2429
4	722	16.5	910	AA193984	Human neuroblastom
5	694.4	15.9	893	AA194300	Human neuroblastom
6	670	15.3	903	AA194300	Human neuroblastom
7	645.8	14.8	904	AA193983	Human neuroblastom
8	551.6	12.6	615	ABK62818	Rat sequence diffe
9	510.2	11.7	542	ABK39516	DNA encoding lung

10	403.2	9.2	451	16	AA222113	Human gene signatu
11	380.2	8.7	2887	21	AA224879	Human ORF2429
12	302	6.9	317	21	AA225273	Human secreted pro
13	274.8	6.3	590	24	ABK62109	Rat sequence diffe
14	176.2	4.0	196	22	AAH36707	Human colon cancer
15	134.8	3.1	211	23	ABV56446	Human prostate exp
16	107.2	2.5	565	24	ABK62830	Rat sequence diffe
17	79.2	1.8	114955	20	AAK53491	Human adenosine AI
18	78	1.8	114955	20	AAK53491	Human adenosine AI
19	72	1.6	1127	21	AAH24477	Human colon cancer
20	70	1.6	801	20	AAV73801	KSHV LDR terminal
21	68.6	1.6	612	22	AAH71471	Human cervical can
22	67.2	1.5	936	18	AAH95102	Exon 3 encoding am
23	64.8	1.5	2561	22	AAH26500	Rabbit low density
24	64.6	1.5	1614	22	AAH26499	Human low density
25	64.6	1.5	12425	22	AAH26495	Human low density
26	63.2	1.4	1614	22	AAH26499	Human low density
27	63.2	1.4	12425	22	AAH26495	Human low density
28	62.8	1.4	556	23	ABV40063	Human prostate exp
29	62.8	1.4	556	23	ABV40163	Human prostate exp
30	62.8	1.4	556	23	ABV42105	Human prostate exp
31	62.8	1.4	556	23	ABV43601	Human prostate exp
32	62	1.4	801	21	AAA30291	Rhadinovirus cis-
33	62	1.4	801	24	ABA33488	Kaposi's sarcoma-a
34	61.4	1.4	372	23	ABV37528	Human prostate exp
35	61.4	1.4	2307	24	ABK32842	DNA encoding human
36	61.2	1.4	266	23	ABV07596	Human prostate exp
37	61.2	1.4	4411529	22	AA199682	Mycobacterium tube
38	60	1.4	60	24	ABN49572	Human spliced tran
39	59.6	1.4	309	23	ABV44994	Human prostate exp
40	59.6	1.4	626	23	ABV06941	Human prostate exp
41	59.6	1.4	109519	22	AA508693	Microsomopora DNA
42	59.2	1.4	309	23	ABV44994	Human prostate exp
43	58.8	1.3	2388	22	AAE52196	DNA encoding memb
44	58.8	1.3	2388	24	ABK12568	DNA encoding cell
45	58.6	1.3	989	21	AAA02539	Human colon cancer

ALIGNMENTS

RESULT 1	
AA224879	AA224879 standard; DNA; 1509 BP.
ID	AA224879; standard; DNA; 1509 BP.
XX	
AC	AA224879;
XX	
DT	02-DEC-1999 (first entry)
XX	
DE	Human secreted protein gene 69 clone HCEK38.
XX	
KW	Human; secreted protein; fusion protein; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
OS	Homo sapiens.
XX	
PN	WO9947540-A1.
XX	
PD	23-SEP-1999.
XX	
PF	18-MAR-1999; 99WO-US05804.
XX	
PR	19-MAR-1998; 98US-0078563.
PR	19-MAR-1998; 98US-0078566.
PR	19-MAR-1998; 98US-0078573.
PR	19-MAR-1998; 98US-0078574.
PR	19-MAR-1998; 98US-0078576.

PR 19-MAR-1998: 9805-0078577.
PR 19-MAR-1998: 9805-0078578.
PR 19-MAR-1998: 9805-0078579.
PR 19-MAR-1998: 9805-0078581.
PR 01-APR-1998: 9805-0080312.
PR 01-APR-1998: 9805-0080313.
PR 01-APR-1998: 9805-0080314.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR:
PI Wei Y, Endress GA, Duan KD, Kyaw H, Edner R, Lafleur DW:
PI Oisen HS, Shi Y, Moore PA:
XX MPI: 1999-562050/47.
XX P-PSDB: AAY41376.
PT New isolated human genes, useful for diagnosis and treatment of e.g.
PT cancers, neurological disorders, immune diseases, inflammation or blood
PT disorders
XX
XX Claim 1: Page 340: 484pp: English.

XX
XX This sequence represents a nucleic acid molecule which encodes a
XX secreted human protein. The gene number, and the clone it is derived
XX from, are detailed in the descriptor line. The gene can be used to
XX generate fusion proteins by linking to the gene to a human immunoglobulin
XX Fc portion (e.g. AA224802) for increasing the stability of the fused
XX protein as compared to the human protein only.
XX The invention relates to 95 novel genes and their fragments (nucleic
XX acid sequences: AA224811-224907; amino acid sequences AA1308-V41404)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 95
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AA224811 for described uses).

Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 other;

Query Match 33 8%; Score 1478; DB 20; Length 1509;
Best Local Similarity 99.98; Pred. No. 3.1e-291;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 2859 GATGACTAATGAGCTTCTCATCTTGTAAATAATTAATTTGATGTGACATCTT 2918
DB 9 GATGACTAATGAGCTTCTCATCTTGTAAATAATTAATTTGATGTGACATCTT 68
QY 2919 GGTCTCTCCCTCCCGTTTGTAAATATCAGATAGCATCTCCAGGCCACTTTGTC 2978
DB 69 GGTCTCTCCCTCCCGTTTGTAAATATCAGATAGCATCTCCAGGCCACTTTGTC 128
QY 2979 TCAGGTAAAGATCCCTATTACATCTGAAAGAAATAGAGCAAGACCTCGTCTCA 3038
DB 129 TCAGGTAAAGATCCCTATTACATCTGAAAGAAATAGAGCAAGACCTCGTCTCA 188
QY 3039 AATATATAGGAATGCTTCTTCTTACCTTCAGAGCATTTGTGTGAACAAGTAGGGGT 3098
DB 189 AATATATAGGAATGCTTCTTCTTACCTTCAGAGCATTTGTGTGAACAAGTAGGGGT 248
QY 3099 CTAATCTCTAGAGTAGGGGCTTTATCTTAAAGAGAAATATGTCCTCCAGATTATTAG 3158
DB 249 CTAATCTCTAGAGTAGGGGCTTTATCTTAAAGAGAAATATGTCCTCCAGATTATTAG 307
QY 3159 CACTTTAGAGGAAAGCAAGGTATGTA-GGGGTGTGGCTGGGCCATCAGTGGAGCAC 3217
DB 308 CACTTTAGAGGAAAGCAAGGTATGTA-GGGGTGTGGCTGGGCCATCAGTGGAGCAC 367
QY 3218 GAAGAGAGATGGATACCATTTGTGGAGAGAGAAAGTTCTTCAGGGGCTCCCAT 3277
DB 368 GAAGAGAGATGGATACCATTTGTGGAGAGAGAAAGTTCTTCAGGGGCTCCCAT 427

QY 3278 GCTAAAGTTTGTGAGATGTTGATCTGTGCTCTCTGAGATTGACTTTAAAGAAATTA 3337
DB 428 GCTAAAGTTTGTGAGATGTTGATCTGTGCTCTCTGAGATTGACTTTAAAGAAATTA 487
QY 3338 TTCTGGCAGACATGATGATTTCTTGGATGATCTCTGCTTATTCTCTCTTTGTGT 3397
DB 488 TTCTGGCAGACATGATGATTTCTTGGATGATCTCTGCTTATTCTCTCTTTGTGT 547
QY 3398 GT 3457
DB 548 GT 607
QY 3458 GGGCTCTCTAATAGGAGAACCTGCTGTAACTTATTCAGACAGAGATGTAGAGAAATA 3517
DB 608 GGGCTCTCTAATAGGAGAACCTGCTGTAACTTATTCAGACAGAGATGTAGAGAAATA 667
QY 3518 GCACTTAATTCACATAGGGGCTCTCATCTCACACCTTAAGAGAGAGATTTGAAAGAA 3577
DB 668 GCACTTAATTCACATAGGGGCTCTCATCTCACACCTTAAGAGAGAGATTTGAAAGAA 727
QY 3578 TGAGCCAGATTTCTTGTCTTCATCATTTTAATGTGGCAGAGCTGTCAGTTTCTTAC 3637
DB 728 TGAGCCAGATTTCTTGTCTTCATCATTTTAATGTGGCAGAGCTGTCAGTTTCTTAC 787
QY 3638 TCTTACCTATGTATATTTCTTCTGATACGTGTCCAAAAGAAAAAGACCAATCAGTGT 3697
DB 788 TCTTACCTATGTATATTTCTTCTGATACGTGTCCAAAAGAAAAAGACCAATCAGTGT 847
QY 3698 CTTCTGACCTTTGTTCTTATCCCTGATGTTCTTCTGATGATGATGATGATGATGATG 3757
DB 848 CTTCTGACCTTTGTTCTTATCCCTGATGTTCTTCTGATGATGATGATGATGATGATG 907
QY 3758 TAATTTGGGATAGATTAAGCAATTAACCAATGTTGTTGCTTACCCAGAGGAGCTC 3817
DB 908 TAATTTGGGATAGATTAAGCAATTAACCAATGTTGTTGCTTACCCAGAGGAGCTC 967
QY 3818 CCGAGTTTGTACTTAAGTACGATGAGAGAAATCAGAGAGTGTATGTGCGCAGATT 3877
DB 968 CCGAGTTTGTACTTAAGTACGATGAGAGAAATCAGAGAGTGTATGTGCGCAGATT 1027
QY 3878 AAGTATATCTAATTTCTTCTGTTCTCCCTGCTCCAGAGACCTTATTTATTTGCCCC 3937
DB 1028 AAGTATATCTAATTTCTTCTGTTCTCCCTGCTCCAGAGACCTTATTTATTTGCCCC 1087
QY 3938 TCTTACGATTAATTTCTTCTGATTTGATTTGATGAGAGAGAGTGTGACAGATTA 3997
DB 1088 TCTTACGATTAATTTCTTCTGATTTGATTTGATGAGAGAGAGTGTGACAGATTA 1147
QY 3998 AGCAAGTTTCAAGTCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATTAAGTATT 4057
DB 1148 AGCAAGTTTCAAGTCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATTAAGTATT 1207
QY 4058 TTTTCCCTACATGGAGTACACAGTGTGAATTCATCTCACTGAAGCCCTGCAGTTC 4117
DB 1208 TTTTCCCTACATGGAGTACACAGTGTGAATTCATCTCACTGAAGCCCTGCAGTTC 1267
QY 4118 TCTTAAACATAGTGTGTTGTTTCTTTTAAACAAAGTTTAAGCTAGTTAAATTA 4177
DB 1268 TCTTAAACATAGTGTGTTTCTTTTAAACAAAGTTTAAGCTAGTTAAATTA 1327
QY 4178 AAAAAATCTGCTGTCTGCTACTTCACTGCTTTGTTTATGCCATTTCAATATGTTGTCG 4237
DB 1328 AAAAAATCTGCTGTCTGCTACTTCACTGCTTTGTTTATGCCATTTCAATATGTTGTCG 1387
QY 4238 TGTGTAAATTCATTAATCTTGTATACATTTGATGTGTAAATTTGTTGTTGTTGTTG 4297
DB 1388 TGTGTAAATTCATTAATCTTGTATACATTTGATGTGTAAATTTGTTGTTGTTGTTG 1447
QY 4298 ATCTTAATAAGATTCATTTGTAATTAACATATGCTGTGTTAAAAAAGAAAAA 4357
DB 1448 ATCTTAATAAGATTCATTTGTAATTAACATATGCTGTGTTAAAAAAGAAAAA 1507
QY 4358 AA 4359

KW	severe combined immunodeficiency, malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinketsu RA, Leach M;
DR	WPI; 2000-602362/57.
XX	
PT	P-PSDB; AAB42665.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
XX	useful for treating e.g. cancers, proliferative disorders,
XX	neurodegenerative disorders and cardiovascular disease -
PS	Claim 5; Page 4044-4045; 5507ep; English.
XX	
CC	AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC	antiproliferic; antiparkinsonian; nootropic; neuroprotective;
CC	osteoprotic; anticonvulsant; antiarthritic; immunosuppressant;
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	antithyroid; and antihaemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hypernatensio, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SQ	Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;
XX	
QY	Query Match 18.0%; Score 785.4; DB 21; Length 2887;
XX	Best Local Similarity 99.7%; Pred. No. 4.2e-150;
XX	Matches 797; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
DB	655 CGGGCAGGAGGAGGAGATGTGAAGTAAGTAGAGAGAGCCCTCTGGGGGTGGTGGCGCT 714
XX	
DB	3 CGGGCCGGGAGGAGGAGATGTGAAGTAAGTAAGTAAGCCCTCTCGGGGTGGTGGCGCT 62
XX	
OY	CGGGAGCCTCCAGTCAGGCGCCCTGCGCCCAAAAGATAGCTCTGCTGCAATTGGACCTCA 774
XX	
DB	63 CGGGAG-CTCCAGTCAGGCGCCCTGCTCAAAACAGATCTCTTCTGCTGCAATTGGACCTCA 121
XX	
OY	775 TCGAAGACAGCAGCAGCAGCTGCGAGGCCAAGAAAGAGATCGAGGAGCTGAAGTCAG 834
XX	
DB	122 TCGAAGACAGCAGCAGCAGCTGCGAGGCCAAGAAAGAGATCGAGGAGCTGAAGTCAG 181
XX	
OY	835 AGAGAGACAGCCTCTTCTGCTGAGTTGAGCTATGTGAAGGCGGATGAGCTGTGAAGA 894
XX	

Db	182	AGAGAGACACCGCTCCTTGTGCGATTGAACGTAATGGAAGGCGGATGACAGTGGTAAACA	241
Qy	895	AGGATAACGAGAAAGAAAGGACACACAGCTGTTTCAGGGCTATGAACATGAAGAGAGAGG	954
Db	242	AGGATAACGAGAAAGAAAGGACACAGCTGTTTCAGGGCTATGAACATGAAGAGAGAGG	301
Qy	955	AAACAGAGCTATCTGAGAAATTTAACTGGAGTGGCCAGCCGGAGCTTTCGAGACATCCC	101
Db	302	AAACAGAGCTATCTGAGAAATTTAACTGGAGTGGCCAGCCGGAGCTTTCGAGACATCCC	361
Qy	1015	AGACCTGCGCTCCCAAGCCCTTCTCATGTGGGGGAGTGGAAAGGAGACATAAAGGAAT	107
Db	362	AGACCTGCGCTCCCAAGCCCTTCTCATGTGGGGGAGTGGAAAGGAGACATAAAGGAAT	421
Qy	1075	CCCCATTTGGAGTACAGAAAGAAAGACTCTGTTTAAAAAGCTGGCTCTGAATTTTCAA	113
Db	422	CCCCATTTGGAGTACAGAAAGAAAGACTCTGTTTAAAAAGCTGGCTCTGAATTTTCAA	481
Qy	1135	AAGTCAAAACAAAACCTCTTAAGCAGCTGCTCTTAAAGAGAAACCTGTGGTTCTTAT	119
Db	482	AAGTCAAAACAAAACCTCTTAAGCAGCTGCTCTTAAAGAGAAACCTGTGGTTCTTAT	541
Qy	1195	CTGAACCTGTTTGTAAAGCTGAATTAGGAGGCCAAGAAACCCAGAAAAAGCCCGGCTT	125
Db	542	CTGAACCTGTTTGTAAAGCTGAATTAGGAGGCCAAGAAACCCAGAAAAAGCCCGGCTT	601
Qy	1255	CAGTGGACACCCACCAAGACTCTCCACTCCCCAAAAGGAGACCCAGACCCATCCCAAG	131
Db	602	CAGTGGACACCCACCAAGACTCTCCACTCCCCAAAAGGAGACCCAGACCCATCCCAAG	661
Qy	1315	AGAAAGCCTTCTCAAGTGAATRGAAGATTTTGCCTTCTTCCACACAGAAATGTATT	137
Db	662	AGAAAGCCTTCTCAAGTGAATRGAAGATTTTGCCTTCTTCCACACAGAAATGTATT	721
Qy	1375	TGTGTGCTGGGACACGACCTCCGCCATCAACCGTTACATTACGGGAATCCTCTCCAAAGA	143
Db	722	TGTGTGCTGGGACACGACCTCCGCCATCAACCGTTACATTACGGGAATCCTCTCCAAAGA	781
Qy	1435	AGGAGGAGACTGTAGCAAG	1453
Db	782	AGGAGGAGACTGTAGCAAG	800
RESULT 4			
ID	AAI93984/c	standard; cDNA; 910 BP.	
XX	AAI93984;		
XX	AC		
XX	DT	13-NOV-2001 (first entry)	
XX	DE	Human neuroblastoma expressed polynucleotide SEQ ID NO 59.	
XX	KM	Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200166719-A1.	
XX	PD	13-SEP-2001.	
XX	PF	02-MAR-2001; 2001WO-JP01629.	
XX	PR	07-MAR-2000; 2000JP-0159195.	
XX	PA	(CHTB-) CHITBA PREFECTURE.	
XX	PA	(HISM) HISAMITSU PHARM CO. LTD.	
XX	PI	Nakagawara A;	
XX	DR	WPI; 2001-565584/63.	
XX	Nucleic acids originating in gene expressed in human neuroblastoma,		

Db 645 ATTTCAGCATGTGTC-GGTTCTTAATTTGGTATGATGAGCAAAATTTACCATGTGT 587
 Qy 3796 TTGTGCCCTACCCAGGGGACTCCCGAGTTCTCAGCTTGAAGTGAAGTGAAGAAATCCAC 3855
 Db 586 TTGTGCCCTACCCAGGGGACTCCCGAGTTCTCAGCTTGAAGTGAAGTGAAGAAATCCAC 527
 Qy 3856 -GAGGTGCTATTC-TGGCCAGATTTAAGTATCTATTTCTGTTCTCCCTCCCTC 3913
 Db 526 GAGAGTCTATCTTCCGCAATTTAAGTATCTATTTCTGTTCTCCCTCCCTC 467
 Qy 3914 AGAGCTCTATTTATTTATGTCCTCTCTAGTAAATTCCTTTGATTTGACTTTGTT 3973
 Db 466 AGGACCTCTATTTATTTATGTCCTCTCTAGTAAATTCCTTTGATTTGACTTTGTT 407
 Qy 3974 GAGAGGAGGTTGGACGATGATTTAGCAAGTTCGCAAAATTTACAGTGTGTTAGA 4033
 Db 406 GAGAGGAGGTTGGACGATGATTTAGCAAGTTCGCAAAATTTACAGTGTGTTAGA 347
 Qy 4034 GTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGATACAACTGTGAATTCAT 4093
 Db 346 GTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGATACAACTGTGAATTCAT 287
 Qy 4094 CTTCACATGAAGGCCCTGCGAGTCTCTAAACATAGTGTGTTGTTTCTTTAAACAAG 4153
 Db 286 CTTCACATGAAGGCCCTGCGAGTCTCTAAACATAGTGTGTTGTTTCTTTAAACAAG 227
 Qy 4154 TTTTAACTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4213
 Db 226 TTTTAACTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 167
 Qy 4214 ATGCCATTTCAATATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4273
 Db 166 ATGCCATTTCAATATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 107
 Qy 4274 TGTAAATTTGCTGT 4333
 Db 106 TGTAAATTTGCTGT 47
 Qy 4334 GGCTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4362
 Db 46 GGCTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18
 RESULT 6
 AA194299
 ID AA194299 standard; cdna; 903 BP.
 AC AA194299;
 DT 13-NOV-2001 (first entry)
 DE Human neuroblastoma expressed polynucleotide SEQ ID NO 374.
 KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
 OS Homo sapiens.
 PN WO20016719-A1.
 PD 13-SEP-2001.
 PF 02-MAR-2001; 2001WO-JP01629.
 PR 07-MAR-2000; 2000JP-0159195.
 PA (CHIB-) CHIBA PREFECTURE.
 PI (HISM) HISAMITSU PHARM CO LTD.
 DR Nakagawara A;
 WP1: 2001-565584/63.

PT Nucleic acids originating in gene expressed in human neuroblastoma,
 PT useful as probe or primer in diagnosing prognosis of human
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
 PT for anti-cancer agents
 PS Claim 1; Page 317; 2979pp; Japanese.
 CC The invention relates to novel genes (AA193926-AA197963) expressed in
 CC human neuroblastoma. The nucleic acids are applicable as a probe or
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
 CC susceptibility indicators or tumour markers for anti-cancer agents. The
 CC gene information for diagnosing prognosis is related to factors similar
 CC to that for N-myc and TrkA genes.
 SQ Sequence 903 BP; 257 A; 169 C; 197 G; 250 T; 30 other;
 Query Match 15.3%; Score 670; DB 22; Length 903;
 Best Local Similarity 93.2%; Pred. No. 9,2e-127;
 Matches 780; Conservative 0; Mismatches 48; Indels 9; Gaps 8;
 Qy 1830 CAGAAATTTGAGCTACCTGCTGTGTGATGAGCGGTAGCCGATGACGATTT-GGAGATCCAGAA 1888
 Db 69 CAGAAATTTGAGCTACCTGCTGTGTGATGAGCGGTAGCCGATGACGATTTGGGAGATCCAGAA 128
 Qy 1889 GAAGCAAAACCTCCACCGGAGCTGTGGAATAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1948
 Db 129 GAAGCAAAACCTCCACCGGAGCTGTGGAATAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 188
 Qy 1949 GATAGT 2008
 Db 189 GATAGT 248
 Qy 2009 CACATATGTATACATGCTGT 2068
 Db 249 CACATATGTATACATGCTGT 308
 Qy 2069 CTCTGATTTACAAAACCTTTTCATTCGCTTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2128
 Db 309 CTCTGATTTACAAAACCTTTTCATTCGCTTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 368
 Qy 2129 TTTCTTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATGGC-AGATTTC 2187
 Db 369 TTTCTTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTC 428
 Qy 2188 AGAAGTTTCAAGGGCTGT 2247
 Db 429 AGAAGTTTCAAGGGCTGT 488
 Qy 2248 TTTCAATAGATGTGGAAGATGAAGCAAGCCCTTCTTATAGACTGTGCTGATGGCAC 2307
 Db 489 TTTCAATAGATGTGGAAGATGAAGCAAGCCCTTCTTATAGACTGTGCTGATGGCAC 548
 Qy 2308 TCTTCTACCCCTGT 2366
 Db 549 TCTTCTACCCCTGT 608
 Qy 2367 AAACAAACCACTTCACCATGAGCTTTAGAGCC-AGAAGAGGAATGAC-AAAGCAAGCAT 2424
 Db 609 AAACAAACCACTTCACCATGAGCTTTAGAGCC-AGAAGAGGAATGAC-AAAGCAAGCAT 668
 Qy 2425 GAAGCAAGCATTTTCACA--GAGTAGAAAGACATCGAGAGTGTGTGTGTGTGTGTGTGTGT 2482
 Db 669 GAAGCAAGCATTTTCACA--GAGTAGAAAGACATCGAGAGTGTGTGTGTGTGTGTGTGTGT 728
 Qy 2483 GAAAGATAGT 2542
 Db 729 GAAAGATAGT 786
 Qy 2543 GGCTGTGATTTATTTTACATTAAGATGAAGCAATTTCTTACTGCTCAAAAACCAAAAT 2602
 Db 787 GGCTGTGATTTATTTTACANTTAAGGATACAAANTTTCTTAACNGCAAAACCAAAATC 846
 Qy 2603 CTTTGGAAAAGAAAGTGGGATGT 2659

Db 847 CTTGGGAAAAGAAATTTGGGATGNTANTTCAAAACAGTTACAGCTGTTACAAA 903

RESULT 7

AA193983 standard; cDNA; 904 BP.

AA193983;

13-NOV-2001 (first entry)

Human neuroblastoma expressed polynucleotide SEQ ID NO 58.

Human neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

Homo sapiens.

MO20016719-A1.

13-SEP-2001.

02-MAR-2001; 2001WO-JP01629.

07-MAR-2000; 2000JP-0159195.

(CHIB-) CHIBA PREFECTURE.

(HISM) HISAMITSU PHARM CO LTD.

Nakagawara A;

WPI; 2001-565584/63.

Nucleic acids originating in gene expressed in human neuroblastoma.

useful as probe or primer in diagnosing prognosis of human neuroblastoma, malignancy and susceptibility indicator or tumour marker for anti-cancer agents

Claim 1; Page 78; 2979pp; Japanese.

The invention relates to novel genes (AA193926-AA197963) expressed in human neuroblastoma. The nucleic acids are applicable as a probe or primer in diagnosing the prognosis of human neuroblastoma, malignancy and susceptibility indicators or tumour markers for anti-cancer agents. The gene information for diagnosing prognosis is related to factors similar to that for N-myc and TrkA genes.

Sequence 904 BP; 265 A; 163 C; 198 G; 244 T; 34 other;

Query Match 14.8%; Score 645.8; DB 22; Length 904;

Best Local Similarity 93.0%; Pred. No. 7.8e-122;

Matches 736; Conservative 0; Mismatches 44; Indels 11; Gaps 6;

1834 ATTTGAGCTACCTGTTGGATGAGCGTACCGATCCAGATTTGAGATCCAGAGAAGC 1893

45 AATTGGCTACCTGTTGGATGAGCGTACCGATCCAGATTTGAGATCCAGAGAAGC 104

1894 AACACCTCACCGGAGGTGTAAGAAATAGCTGTGCGAAGAACCTGCTTCAAGTAG 1953

105 AACACCTCACCGGAGGTGTAAGAAATAGCTGTGCGAAGAACCTGCTTCAAGTAG 164

1954 TTGTAGCATGCAATTTCCGAGAGTGGCAGAGACCTGTATATGACCTTTGCTTCAT 2013

165 TTGTAGCATGCAATTTCCGAGAGTGGCAGAGACCTGTATATGACCTTTGCTTCAT 224

2014 ATGTATACATCGTGTATATATACCTTTTCACTTCTTCACTTGTGTTTCACTCTG 2073

225 ATGTATACATCGTGTATATATACCTTTTCACTTCTTCACTTGTGTTTCACTCTG 284

2074 ATTTCACAAAACCTTTCACTTGTGTTTCACTTGTGTTTCACTTGTGTTTCACTCTG 2133

285 ATTTCACAAAACCTTTCACTTGTGTTTCACTTGTGTTTCACTTGTGTTTCACTCTG 344

QY 2134 TTTCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAGT 2193

Db 345 TTTCCTTTTGGGAAATGGGCTCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAGT 404

QY 2194 TTTCAGGGCTCTGTTTCTATACATTTGCTTATGTTAAAGGGGTAAGGGCTCTCTCAAT 2253

Db 405 TTTCAGGGCTCTGTTTCTATACATTTGCTTATGTTAAAGGGGTAAGGGCTCTCTCAAT 464

QY 2254 AGACATGTGGAAGATGAAGCAGCCCTTCTTTAGACTGTGCTGCATGGCAGCTCTCT 2313

Db 465 AGACATGTGGAAGATGAAGCAGCCCTTCTTTAGAACTGTGCTGCATGGCAGCTCTCT 524

QY 2314 CACCTT-GGTACACCTCTCTTATAGTGGTAT-AGTATTTTAACTTAAATAAACA 2371

Db 525 CACCTTGGTACACCTCTCTTATAGTGGTATAGTATTTTAACTTAAATAAACA 584

QY 2372 AACACCTCACCGGAGGTGTAAGAAATAGCTGTGCGAAGAACCTGCTTCAAGTAG 2431

Db 585 AACACCTCACCGGAGGTGTAAGAAATAGCTGTGCGAAGAACCTGCTTCAAGTAG 644

QY 2432 GCCATCTTCAC-AGATGAAAGAAAGACATCGAGAGTGTGATTAAGTCTGAAAGAT 2490

Db 645 GCCATCTTCACAGAGATGAAGAAAGACATCGAGAGTGTGATTAAGTCTGAAAGAT 704

QY 2491 AGTGTTCATTTGAAACTATTTCTGTATACAGTATGATGGAGAGATTTGGCTGTA 2550

Db 705 AATTGTCAAT--GAAACAATCTGTATACATCAAGTGAAGAAAGATTTGGCGNGA 762

QY 2551 TTATTTTTCAGTTAATGATTAACAATTTCTTACTGCTCAAAAACCAAAATTTTGGAA 2610

Db 763 TAA--TTTTCAGTAAAGGNAATAATCTTAAGNT---CAAAACAAAACCTTGGAA 816

QY 2611 AAGAAAGTGGG 2621

Db 817 AANAAATGGG 827

RESULT 8

ABK62818 standard; cDNA; 615 BP.

ABK62818;

18-JUN-2002 (first entry)

Rat sequence differentially expressed in response to a hepatotoxin #725.

Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;

differential expression; centrilobular necrosis; steatosis.

Rattus norvegicus.

WO200210453-A2.

07-FEB-2002.

30-JUL-2001; 2001WO-US23872.

31-JUL-2000; 2000US-222040P.

02-NOV-2000; 2000US-244880P.

11-MAY-2001; 2001US-290023P.

15-MAY-2001; 2001US-290645P.

22-MAY-2001; 2001US-292336P.

06-JUN-2001; 2001US-295798P.

13-JUN-2001; 2001US-297457P.

19-JUN-2001; 2001US-298884P.

09-JUL-2001; 2001US-303459P.

(GENE-) GENE LOGIC INC.

Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

WPI; 2002-241625/29.

XX Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or
 PT cells exposed to the toxin and comparing these to gene expression in
 PT unexposed tissues or cells -

XX Claim 1; Seq ID No 725; 239pp; English.

PS The invention relates to methods for predicting toxic effects of
 CC compounds or the progression of these toxic effects by determining the
 CC global changes in gene expression in tissues or cells exposed to the
 CC toxin and comparing these to gene expression in unexposed tissues or
 CC cells. Also included are methods of predicting at least one toxic
 CC effect of a compound or progression of a toxic effect, preferably the
 CC hepatotoxicity of a compound, comprising detecting the level of
 CC expression in a tissue or cell sample exposed to the compound of two or
 CC more genes listed in the specification, where differential expression of
 CC the genes is indicative of at least one toxic effect or progression.
 CC The method can also be used to identify an agent which modulates the
 CC toxic response and predict cellular pathways that a compound modulates
 CC in a cell. The methods utilise a set of at least two probes (on a solid
 CC support in kit form), where each of the probes comprises a sequence that
 CC specifically hybridises to a gene listed in the specification, a computer
 CC system comprising a database containing information identifying the
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
 CC set of genes comprising at least two genes listed in the specification,
 CC and a user interface to view the information used to present information
 CC identifying the expression level in a tissue or cell of at least one gene
 CC listed in the specification. The method is useful for elucidating global
 CC changes in gene expression and for identifying toxicity markers in
 CC tissues or cell exposed to a known toxin. The genes may be used as
 CC toxicity markers in drug screening and toxicity assays. The genes and
 CC gene expression information may be used as diagnostic markers for the
 CC prediction or identification of the physiological state of tissue or cell
 CC sample that has been exposed to a compound or agent. Hepatotoxicity
 CC is characterised by centrilobular necrosis and steatosis. The present
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
 CC which is differentially expressed in response to a hepatotoxic agent.

XX Sequence 615 BP; 84 A; 235 C; 224 G; 71 T; 1 other;

Query Match 12.6%; Score 551.6; DB 24; Length 615;

Best Local Similarity 94.6%; Pred. NO. 1e-102; Indels 3; Gaps 1;

Matches 582; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

XX 112 CGCGCCCTGGCGGCGCAATCTGAGCAGCGACTGAGAGCGGGCTGCGCGCTGG 171
 DB 1 CGGCCCCCTGGCGGCGCAACCCGAGAGAGACTGGACTGACGAGGGGGCTGG 60
 QY 172 GCGGGCCCGAGAGAGAGCTGGGGGCGGCAAGCCCACTTCCCGCGGCAAGC 231
 DB 61 GCGGGCCCGAGAGAGAGCTGGGGGCGGCGCAAGCCCACTTCCCGCGGCAAGC 120
 QY 232 TCAGAGAGCGGGGGCGGCGCTGCTCCAGAGGGGCGGAGCCCGCGGCTTCCCGG 291
 DB 121 TCAGAGAGCGGGGGCGGCGCTGCTCCAGAGGGGCGGAGCCCGCGGCTTCCCGG 180
 QY 292 CCGGCT--GCGGCGCAAGAGCGGCGGCTTGTACTCCGCGCGGGGCGGCGGCG 348
 DB 181 CTGGCTGGCGGCGGCGCAAGGGCGGCGGCTTGTACTCCGCGCGGGGCGGCGGCG 240
 QY 349 AGCAGAGAGAGAGCTGGGGGCGGCTGCTCCGCTGTCGCGCCCGCGGCAAGC 408
 DB 241 AGCAGAGAGAGAGCTGGGGGCGGCTGCTCCGCTGTCGCGCCCGCGGCAAGC 300
 QY 409 AAGCGGCGATGGGGGAGCGGCGGAGCGGAGCGGCTGACGCGCGGCGCAAG 468
 DB 301 AAGCGGCGATGGGGGAGCGGCGGAGCGGAGCGGCTGACGCGCGGCGCAAG 360
 QY 469 ATCAGCGGTGCTGCCATTCACAGCGGCTCTCTCTGCGGGCGGCGCAAGAGAGCTTACGC 528
 DB 361 ATCAGCGGTGCTGCCATTCACAGCGGCTCTCTCTGCGGGCGGCGCAAGAGAGCTTACGC 420

QY 529 CCGGGCTGGGGAGCAAGAGTGCGGCGCCTCCCGGCTGCCACCGCTTGGAGCCGCGG 588
 DB 421 CCGGGCTGGGGAGCAAGAGTGCGGCGCCTCCCGGCTGCCACCGCTTGGAGCCGCGG 480
 QY 589 GACCCCGACACACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 648
 DB 481 GACCCCGACACACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 QY 649 CCGTGGCGGCGAGCGAGGCGAGTGAAGAGTGAAGAGAGCCCTCTCGGGGTGCTG 708
 DB 541 CCGTGGCGGCGAGTGAAGAGGCGAGTGAAGAGTGAAGAGAGCCCTCTCGGGGTGCTG 600
 QY 709 GCGGCTGGGAGCCT 723
 DB 601 GCNGCTCGGGAGCCT 615

RESULT 9

ABK39516
 ID ABK39516 standard; cDNA; 542 BP.

XX ABK39516;

XX 21-MAY-2002 (first entry)

XX DNA encoding lung tumour protein associated contig 228.

XX Lung tumour; cancer; T cell; immune response stimulator;

XX cytostatic; contig; ss.

XX Homo sapiens.

XX W0200204514-A2.

XX 17-JAN-2002.

XX 10-JUL-2001; 2001MO-US22058.

XX 11-JUL-2000; 2000US-0614124.

XX 29-AUG-2000; 2000US-0651563.

XX 08-SEP-2000; 2000US-0658824.

XX 26-SEP-2000; 2000US-0671325.

XX 06-OCT-2000; 2000US-0677419.

XX 30-OCT-2000; 2000US-0702705.

XX 13-DEC-2000; 2000US-0736457.

XX 03-MAY-2001; 2001US-0849626.

XX (CORI-) CORIXA CORP.

XX Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MM;
 PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
 PI Menabbi A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
 XX WPI; 2002-164634/21.

XX Novel polynucleotide encoding a lung tumour polypeptide useful for
 PT stimulating and/or expanding T cells specific for a tumour protein

XX Claim 1; SEQ ID No 1554; 223pp; English.

XX The invention describes an isolated polynucleotide and polypeptide
 CC useful for stimulating and/or expanding T cells specific for a tumour
 CC protein for determining the presence of a cancer in a patient. A
 CC composition containing the polynucleotide and/or polypeptide is useful
 CC for treating a lung cancer in a patient. The polypeptide is useful for
 CC removing tumour cells from a biological sample. The polynucleotide is
 CC also useful as probe or primer to detect the level of mRNA encoding a
 CC tumour protein. This sequence represents a contig of a DNA sequence
 CC encoding a lung tumour associated protein, described in the method of
 CC the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at

AC AAC6874;
 XX 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; nootropic; hepatoprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
 KW antineumatic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disease; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB42665.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5: Page 4044-4045; 5507pp; English.
 XX
 XX AAC7446 to AAC7606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertensive; antineumatic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers;
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;
 Query Match 8.7%; Score 380.2; DB 21; Length 2887;
 Best Local Similarity 95.6%; Pred. No. 1,4e-67;

Matches 391; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 1045 GGGGAGTGGAAAGGAGCAATAAAGAAATCCCATTTGGAGTACAGAAAGAAAGATC 1104
 DB 2493 GGAGAGGGGTGGAGGGGTTAGCTTCTACCCCATTTGGAGTACAGAAAGAAAGATC 2434
 QY 1105 CTGTAAAAAGCTGGCTCTGAAATTTCAAAAGTCAAAACAAAACCTCTAACCTCTC 1164
 DB 2433 CTGTAAAAAGCTGGCTCTGAAATTTCAAAAGTCAAAACAAAACCTCTAACCTCTC 2374
 QY 1165 CTATTAAAGAGAACCTCTGGTCTCTATCTAATCTTTTAAAGTGAATTTAGGA 1224
 DB 2373 CTATTAAAGAGAACCTCTGGTCTCTATCTAATCTTTTAAAGTGAATTTAGGA 2314
 QY 1225 GCCAAGAAACCCAGAAAGCCCGGCTTCACTGAGACCCCAAGACCTCTCAGCTC 1284
 DB 2313 GCCAAGAAACCCAGAAAGCCCGGCTTCACTGAGACCCCAAGACCTCTCAGCTC 2254
 QY 1285 CCCAAAGGAGCCAGCCACCCATCCCAAGAGAAACCTTCTCACTGAGATGAAGATT 1344
 DB 2253 CCCAAAGGAGCCAGCCACCCATCCCAAGAGAAACCTTCTCACTGAGATGAAGATT 2194
 QY 1345 TGCCGTACCTTTCCACACAGAAATGATTTGTGCTGGACCAAGCCCTCCCACTAC 1404
 DB 2193 TGCCGTACCTTTCCACACAGAAATGATTTGTGCTGGACCAAGCCCTCCCACTAC 2134
 QY 1405 CGTTACCATTTACGGAATCCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
 DB 2133 CGTTACCATTTACGGAATCCTCTCCAAAGAGAGAGAGACTGTAGCAAG 2085
 RESULT 12
 AAC25273
 ID AAC25273 standard; cDNA; 317 BP.
 AC AAC25273;
 AC
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 29348.
 DE
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 OS
 XX
 PN EP1033401-A2.
 PD 06-SEP-2000.
 PD
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 PF
 XX
 PR 26-FEB-1999; 99US-0122487.
 PR
 XX
 PA (GEST) GENSET.
 PA
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 PI
 DR WPI; 2000-500381/45.
 DR
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PT
 XX
 PS Claim 1: SEQ ID 29348; 71pp + CD-ROM; English.
 XX
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences

Db 197 TTAACAAGTTTAAGCTAGTGTAAATAAATAAGAGAAGAGCTGTCCTCCACCTC 138
QY 4203 AGCTTGTGTTTAAAGCCATTTCATATGTTGCTGTGTTGTAATTCATAACTTTGATAC 4262
Db 137 AGCTTGTGTTTAAAGCCATTTCATATGTTGCTGTGTTGTAATTCAT-ACITTTGATAC 79
QY 4263 CATTTCTGATGCTAAATATGTTGCTGTGTAATATCTTATTAAGAGTTCAATTTGATA 4322
Db 78 CATTTCTGATGCTAAAGCTGCTGTGTTGTAATATCTTATTAAGAGTTCAATTTGATA 19
QY 4323 TAAACTATGTGCTGCTT 4340
Db 18 TAAACTATGTGCTGCTT 1

RESULT 14

AAH36707 standard; cDNA; 196 BP.

AAH36707;

03-SEP-2001 (first entry)

Human colon cancer antigen encoding cDNA SEQ ID NO:3789.

Human; colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma, ss.

Homo sapiens.

WO200122920-A2.

05-APR-2001.

28-SEP-2000; 2000MO-US26524.

28-SEP-1999; 9905-0157137.

03-NOV-1999; 9905-0163280.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA.

WPI; 2001-235357/24.

P-PSDB; AAG77300.

Claim 1; Page 5645-5647; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated P, by inserting the nucleic acids into a host cell and culturing the cell, to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for CC SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 196 BP; 39 A; 42 C; 42 G; 62 T; 11 other;

Query Match 4.0%; Score 176.2; DB 22; Length 196;
Best Local Similarity 93.2%; Pred. No. 2e-26;
Matches 178; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3810 GGGGACTCCCACTTCTGACTGAGTGAAGTGAAGAACAAATCCAGAGTGTATCTG 3869
Db 5 GGNNTTTCCTCCAGTTCTGACTGAGTGAAGTGAAGAACAAATCCAGAGTGTATCTG 64
QY 3870 CCAGATTTAAGTGAATTCATTTCCCTGCTCCCTCCCTGAGAGCTGTATTTTA 3929
Db 65 CCAGATTTAAGTGAATTCATTTCCCTGCTCCCTCCCTGAGAGCTGTATTTTA 124
QY 3930 TTGTCCTCCCTCTCTGAGTGAATTCCTTTGATTTGAGTGAAGAGAGTGTGAC 3989
Db 125 TTGTCCTCCCTCTCTGAGTGAATTCCTTTGATTTGAGTGAAGAGAGTGTGAC 184
QY 3990 AGTAGATTAGC 4000
Db 185 AGTAGATTAGC 195

RESULT 15

ABV56446 standard; cDNA; 211 BP.

ABV56446;

17-SEP-2002 (first entry)

Human prostate expression marker cDNA 56437.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

Homo sapiens.

WO200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001MO-US05171.

17-FEB-2000; 2000US-183319P.

16-MAR-2000; 2000US-189862P.

25-MAY-2000; 2000US-207454P.

09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.

13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE.

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 10890; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (1) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate, cell carcinogenic potential of a compound;

(g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

SO Sequence 211 BP; 89 A; 37 C; 29 G; 55 T; 1 other;

Query Match 3.1%; Score 134.8; DB 23; Length 211;

Best Local Similarity 98.6%; Pred. No. 5.6e-18;

Matches 136; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4235 CTGCTGTGTAATTCATTAACCTTTGATACCATTTCTGATGCTAAATTTGGTTGCTTGTGA 4294

Db 48 CGGGGTTGTATTCATTAACCTTTGATACCATTTCTGATGCTAAATTTGGTTGCTTGTGA 107

QY 4295 AATATCTTAAAGAGTTCATGTAATAAATATGCTGCTTAAATTTGGTTGCTTGTGA 4354

Db 108 AATATCTTAAAGAGTTCATGTAATAAATATGCTGCTTAAATTTGGTTGCTTGTGA 167

QY 4355 AAAAAAAAAAAAAAAAAA 4372

Db 168 AAAAAAAAAAAAAAAAAA 185

Search completed: July 14, 2003, 12:04:01
Job time : 871 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:11 ; Search time 43 Seconds
(without alignments)
1372.710 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3338

Sequence: 1 MTRSAVFKAAPAGGNPE.....RRCRLFIQKKQPHRTCRK 614

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211	6.5	955	2	male-specific leth
2	202	6.2	854	2	neurofilament trip
3	200	6.2	3938	2	Bassoon protein -
4	199	6.1	1020	1	neurofilament trip
5	197.5	6.1	1072	1	neurofilament trip
6	197.5	6.1	3942	2	Bassoon protein -
7	192	5.9	3534	2	legume protein 2
8	190	5.9	1087	1	neurofilament trip
9	190	5.9	1098	2	probable transcript
10	189.5	5.9	606	2	neurofilament trip
11	189	5.8	539	2	hypothetical prote
12	188	5.8	2715	2	eyelid - fruit fly
13	186.5	5.8	460	2	hypothetical prote
14	184.5	5.7	1520	2	hypothetical prote
15	183.5	5.7	1520	2	hypothetical prote
16	182	5.6	1216	2	hypothetical prote
17	181.5	5.6	2187	2	nascent polypeptid
18	178	5.5	1585	2	hypothetical prote
19	176.5	5.5	7962	2	elastic titin - hu
20	173	5.3	351	1	citrusporozite p
21	173	5.3	1207	2	KIA0641 protein -
22	173	5.3	2022	2	glucan 1,4-alpha-g
23	172.5	5.3	707	2	PTB-associated spl
24	172.5	5.3	1188	2	extensin-like prot
25	172	5.3	1151	2	high molecular mas
26	172	5.3	1870	2	MHC class III hist
27	172	5.3	1870	2	MHC class III hist
28	171.5	5.3	3968	2	trithorax homolog
29	171	5.3	1285	2	ataxin-2 - mouse

30	171	5.3	1647	2	SNF2beta protein -
31	170.5	5.3	418	2	hypothetical prote
32	170	5.3	834	2	hypothetical prote
33	169.5	5.2	858	2	neurofilament trip
34	169.5	5.2	891	2	neurofilament trip
35	169	5.2	268	2	probable prolina-r
36	169	5.2	1613	2	homeotic protein C
37	169	5.2	2342	2	protein BRG1 - hum
38	168.5	5.2	3530	2	hypothetical prote
39	167	5.2	398	2	unconventional myo
40	167	5.2	657	2	hypothetical prote
41	167	5.2	1320	2	probable SF16 prot
42	167	5.2	2142	2	TCOF1 protein - mo
43	166.5	5.1	6642	2	MHC class III hist
44	166	5.1	753	2	protein UNC-89 - C
45	166	5.1	1634	2	Op protein - Kenne
					hypothetical prote

ALIGNMENTS

RESULT 1

S52959 male-specific lethal-1 protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

C:Accession: S52959; S65350

R:Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.

Genetics 134, 545-557, 1993

A:Title: The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes

A:Reference number: S52959; MUID:93314941; PMID:8325488

A:Accession: S52959

A:Molecule type: DNA

A:Residues: 1-955 <PAL>

A:Cross-references: EMBL:L14582

R:Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.

submitted to the EMBL Data Library, April 1993

A:Reference number: S65350

A:Accession: S65350

A:Molecule type: DNA

A:Residues: 1-407, 'S', 409-585, 'M', 587-955 <PAW>

A:Cross-references: EMBL:L14582

C:Genetics:

A:Gene: msl-1

A:Cross-references: FlyBase:FBgn0005617

A:Introns: 329/3

Query Match	6.5%	Score 211;	DB 2;	Length 955;
Best Local Similarity	20.3%	Pred. No. 0.0061;		
Matches	96;	Conservative	91;	Mismatches 171; Indels 114; Gaps 18;
OY	181	PPLATATAGTILASEGMRKSPLOGGGSGSSONACIKOTLL-----LQLDLIE	234	
DB	485	PKHLKAAVAPVKYTKTSRSTLPKA-----NTDIDADAPQKYIANHSGTKOTDVK	537	
OY	235	OOOOLQAKKEI-----ELKSERDT-LIARLERRMQLVKDKNEKRKL	282	
DB	538	TORLOVKTRYEMHDMTGGSPADIKQKWDVSPERKTKSKSILVNDKTTSET	597	
OY	283	FOGYTEEREET---ELSEKIKLECPLESETSOTLP-----PPFGSGRGSK	327	
DB	598	SQSPQDELIDETVRKLAHLKELLSQSHSQVTLKKIRERVAATNLTPP-----	649	
OY	328	GHKRSPPGSTERKTPVKKLAPESKVTYTKPKH---SPIKEPQSGISEYCKRELASQ	384	
DB	650	---SAPVSVT---TTPAPPTSTPTPGSTPOHVTSMQDEISAASKSKAAAOI---	699	
OY	385	ETPEKPRSSVTPPLPLSLNPOK-----GSTPHKEKAFSETEED-----LPLSTP	429	
DB	700	APPLTPQSNSSVSTSTIRKTLNNGSPHTYSKATARGKLGSRRTATFPYSTTWEDQ	759	
OY	430	---EMYLGRMHOPPPSPPLPLRESSPKKETVAVRCIMPSSVAGETSVLAVPSWRDSV	483	

Db 760 EFHCNDEFL-----EEA-----DELLADNPSLEIPKWRDPV 792

QY 484 EPLRDPNSDLLENLDDSVESKRRAKLELDEKRRKMDIORIRORILORLQMLMYKKG 543

Db 793 PPSSDKIDTEL---LSQATERRRHKQYKVDYDKCKDARKMKQIQLRQLMRNDEV 849

QY 544 IQSESP-EVTSFEPDPDVESLMTPLPVVAFGRPLPKLTPQ-NFELPWLD 593

Db 850 LVALDPLRASTFYPLPEDIEAIQFVNEVTVAQFGENVMNBARDFGVPMWD 901

RESULT 2

S02003
neurofilament triplet H protein - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 13-Aug-1999
C:Accession: S02003
R:Breen, K.C.; Robinson, P.A.; Wion, D.; Anderton, B.H.
FEBS Lett. 241, 213-218, 1988
A:Title: Partial sequence of the rat heavy neurofilament polypeptide (NF-H). Identification
A:Reference number: S02003; MUID:89065087; PMID:3143606
A:Accession: S02003
A:Molecule type: mRNA
A:Residues: 1-854 <BRED>
A:Cross-references: EMBL:X13804; NID:957828; PIDN:CAA32038.1; PID:957829
A:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil

Query Match 6.2%; Score 202; DB 2; Length 854;
Best Local Similarity 21.5%; Pred. No. 0.015;
Matches 121; Conservative 92; Mismatches 252; Indels 98; Gaps 22;

QY 11 AAAPAGNPQRDLRYERAALGPEDEPGAAEHFL---LPRHRLKEPGRPLASSQGGSP 67

Db 315 AKSPA-----EVKSPAAYKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 366

QY 68 APSNAGCGGKGRLLPLAGA-----APGQDESMGSGVPLPPEPATQAGIGGE-PA 119

Db 367 AEVKSPEVAKS-----PAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 421

QY 120 AAGAGCSPRRQYAVLPIOTGSLVAAKKEPTPMAGDRGGAASPAATASDPAGPPLPLPG 179

Db 422 EVKSPVYVKSFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFA 471

QY 180 PPLAPATAGTTLAASGRMKSMKSPPLGGGGGGAASQAACLKQILLDLLEQQOQQ 239

Db 472 KSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFA 528

QY 240 LQAKKEIEELKSRDLRLARIEHMERRMOLYKDKNKEKHLFGQYETEREETELSEK 299

Db 529 VKGAKSLAEKSPKSEA-----KSPVKEIKKPAEYKSPKSAKSPKSAKSPKSAKSPKSA 572

QY 300 IKLECPDELSETSLTP--KPFSCGRSGKHKRSPFGSTER-KTPVKKLAPFSVYKT 356

Db 573 AK---SEKAKTLDVKSPEAKP-----PAKEBAKRPALIRSPVOYKSAKEAKSPKEET 625

QY 357 KTRKHSIKPEPCGSLSEYWK---RELRSQETPEKPRSSV-----DTPRLSTP---Q 404

Db 626 RTEVPAKKEEVKSPVEVAKKEPPKVEEEKTPATKTEYKESKKDEAKQKPAEE 685

QY 405 KGPSTH-PKEKASSELEDLYLSTTEMYLCRMHQPPSPPLPLR-ESSPKKEEVAVACLM 462

Db 686 KEPLTEKPKDSPEAKKEE-----AKKKKAAPDEETPAKLVGEEKPEKEDAKAKE 740

QY 463 PSSVA-----GTSVLAIVPSMRDHSVE-----PLRDPNSDLLENLDDSVSKR 506

Db 741 PSKPSSEKPKKEEVPAPEKPKDKTEKTEKSKREKPKMEPRPKRRTAFAKSLANPR 800

QY 507 HAKLEDEKRRKMDIORIREQR 529

Db 801 QKRLKSPLAOTKKTKTASQRRPQR 823

RESULT 3

T42761

Bassoon protein - rat

N:Alternate names: brain-specific synapse-associated protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42761

R:Dieck, S.; Sammarti-Villa, L.; Langeneese, K.; Richter, K.; Kindler, S.; Soyke, A.; W

J. Cell Biol. 142, 499-509, 1998

A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively local

A:Reference number: Z22249; MUID:98345363; PMID:9679147

A:Accession: T42761

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-3938 <DIE>

A:Cross-references: EMBL:Y16563; NID:93413503; PIDN:CAA76287.1; PID:93413504

A:Experimental source: strain Sprague Dawley; brain

A:Function: may be involved in cytomatrix organization at the site of neurotransm

A:Note: component of the presynaptic cytoskeleton

C:Keywords: coiled coil; zinc finger

Query Match 6.2%; Score 200; DB 2; Length 3938;
Best Local Similarity 20.4%; Pred. No. 0.089;
Matches 162; Conservative 56; Mismatches 250; Indels 328; Gaps 32;

QY 3 MRSVAFKAAAPAG---GNPEQRDLRYERAALGPEDE-----PGAFAHFLPRHRK 51

Db 239 LHPSPALPSNAPKOPKPEQERSRSPATQSPRFAAARATSVGPQATAPPEVGR 298

QY 52 LK-----EPGPPLASSOGGSPADSPACG-----GKGRGLLL- 83

Db 299 VSPQPLSTRPSTAEPRPPAGDAQKATTVPSGLAAEOTGGTLGKLGLASLITQA 358

QY 84 -----PAAARQOQESMG-----GSVLPCCPPATKQ----- 111

Db 359 STLKSVQPEADYQGPSPKSGPPKTYVSDASKENGRPPSGGPGTTPAKTEPGRITG 418

QY 112 -----AGIGGP-----AAAGAGCSPR--PKYQAVLPIOTGSLVAAK----- 147

Db 419 PGSGPGLATGTGTPSPKHKRADHQAASKAAAKPKTKPKRAACPLQALNANGSGRAN 478

QY 148 -----EPPPMAGDK-----GGAASPAATASDPAGPPL 175

Db 479 YNTCTACKLVCTLCGFNPPLHVEKTEWMLCLNCOTKRLIEGSLGEPA-----PLPL 530

QY 176 PLPGPPLATATAGTTLAASGRMKSMKSPPLGGGGGSG-----ASSQAACLKQILLDL 230

Db 531 PTPQDEP-----AGVQRAAGASPLKQKGPQGGPQSGSLPPKASFOAKKASP-----QA 580

QY 231 DLLEQOQOQLOAKKEIEELKSRDLRLARIEHMERRMOLYKDKNKEKHLFGQYETEE 290

Db 581 AKASPOKPLPRASP-----PS 596

QY 291 REETELSEK---IKLECPDELSETSLTPPKPSCGSGKHKRSPFGSTERKTPYKTL 347

Db 597 KTSAPAEKKTGTPVKAEP-----VKKP-----PPTAAPPKPAKSGVKRT 639

QY 348 APERSKVKTKTPHPSKKEEP-CGSLSEYVCKRELRSQETPEKPRSSDPPRLSTPQKG 406

Db 640 DP-----ATPVVKKPPEAPKSGEAEPPKRP--TSQDLSRSPQSLSDT-----GVSSDG 686

QY 407 PSTHPRK--KAFSSELEDLYLSTTEMYLCRMHQPPSPPLPLRE-SSPKKEETVARCLMP 463

Db 687 VSSQSEITGVVQVEVQLDSAGVT-----GPRPSPSELHKVSGSMKSPSLAQAVAP 739

QY 464 S---SVAGETSVLAIVPSMRDHSVEPL-----RDPNPSDLLEN 497

Db 740 SGEMSKPPSGSAVEDQKRRPHSLSLIMEAFSDDELGLDILEEDDSLAMGQRQOQDAES 799

QY 498 LDDSVSKRAKLE-----LDEKRRKMDIORIRORILOR- 533

Db 800 SDFGSQLRHDIYEDSSEGLSPLPPOPPARADMTDEEFMRKQIILEKSAEDNLEEDDTA 859

QY 534 LQLRMYKKKGIOESEPEVTSEFPEDDVESIMITPFLPYVAFGRPLPKLTQNFPLMD 593
 DB 860 VSGRGLAKHGAK-----ASARRPRESSQESVALP--- 889
 QY 594 ERSRCRLERIOKKOPPH 609
 DB 890 -----KRRLLPH 895

RESULT 4

OPH0H
 neurofilament triplet H protein - human
 N:Alternate names: neurofilament protein, 112K
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
 C:Accession: S00979
 R:Lees, J.F.; Smeidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
 EMBO J. 7, 1947-1955, 1988
 A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
 A:Reference number: S00979; MUID:88328981; PMID:3138108
 A:Accession: S00979

A:Molecule type: DNA
 A:Residues: 1-1020 <LEE>
 A:Cross-references: EMBL:X15306; NID:g35028; PIDN:CAA33366.1; PID:g1841430
 A:Note: It is uncertain whether Met-1 or Met-2 is the initiator
 C:Genetics:
 A:Gene: GDB:NEFH
 A:Cross-references: GDB:120225; OMTM:162230
 A:Map position: 22q12.1-22q13.1
 A:Introns: 295/1; 361/3; 403/2
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
 F:1-100/Domain: amino-terminal <NTP>
 F:101-410/Domain: rod #status predicted <ROD>
 F:411-1020/Domain: carboxyl-terminal <CTD>
 F:502-826/Region: 14-residue repeats
 F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
 (covalent) #status predicted
 F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 6.1%; Score 199; DB 1; Length 1020;
 Best Local Similarity 23.6%; Pred. No. 0.024;
 Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

QY 90 GQOESMGSGVPLPCPPATKQAGIGGEPAAAGCSPRKYQAVLP-IGSGLVAAKE 148
 DB 489 GEEEAEGEGEETWSP-----AEFAAS-----PEKEAKSPVKEEAKSPAEEKS 532
 QY 149 PTPMAGDKGAASPAAT-----ASDPA-----GPPPLPLPGPPPLPATATAGTLAASGR 198
 DB 533 P-----EKBEAKSPAEVKSPEAKSPAEEKASPEEAKSPAEEKASPEAKSP 587
 QY 199 WKSMSKSPGGGGGSGAASQAACLKQILLDLLEQOQOOLQAKEKEIEELSERDTLL 258
 DB 588 AKKEAKSPAEEKSPAEEKSPVK-----EAKSPAEEKSPAEEKSP----- 629
 QY 259 ARIEMERRMLVKKDNKEKRLKLFQGYETERETELSEKTLKCOPELSTSTQTLPPK 318
 DB 630 AEVKSPEK-----ASPTKEEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPV 684
 QY 319 PFGSGRGKGMKSPGSTERKTVPKTLAPFESVKTTPKHSPIKE-PCGSLSIEWC 377
 DB 685 K---AEAKSPAEEKSPV-KEEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPV 740
 QY 378 KRELRSOE---TPKPRSSVDTPPRLSTPQKGPSTHP---KEKAFSEIEDLPLSTTE 430
 DB 741 KEKASPEKAKSPVAKTLDVKSPEAKTPAKKEEASPADKPEEKASPVKEEVKSPERAK 800
 QY 431 MYLCRMHQPSPPLIRE---SSPKKE---IVARCLMSSVAGETSVLAIVSM---RDHS 482
 DB 801 SPLKADAKAPEKTEPKKEVKSPEKEEKPOEVKKEPKKAAEERKADATPTEKKDSK 860

QY 483 VE--PLRDPNPSDLLENIDDSVFSKRHAKLEDEKRRKRWIDORI--REORITQRLQRM 538
 DB 861 KEAKKKAAPKPKVEKEKEPAVEKKEKVE--AKKEADKKKVPTEKEKAPAVEV-- 916
 QY 539 YKKKGIOESEPEVTSEFPEDDVES 563
 DB 917 -KEDAKPKTEVAK--KEPDAKA 938

RESULT 5

A37221
 neurofilament triplet H protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
 C:Accession: A37221 #25649; A30796; A32757; B25649
 R:Chin, S.S.M.; Liem, R.K.H.
 J. Neurosci. 10, 3714-3726, 1990
 A:Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with
 A:Reference number: A37221; MUID:91038277; PMID:2230956
 A:Accession: A37221

A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1072 <CHI>
 A:Cross-references: GB:AF031879; NID:g2642597; PIDN:AAB87068.1; PID:g2642598
 R:Robinson, P.A.; Wion, D.; Anderton, B.H.
 FEBS Lett. 209, 203-205, 1986
 A:Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
 A:Reference number: A25649; MUID:87080760; PMID:2876828

A:Accession: A25649
 A:Molecule type: mRNA
 A:Residues: 230-318; 472-542 <ROB>
 A:Cross-references: GB:M37227
 R:Dautigny, A.; Pham-Dinh, D.; Rousset, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.
 Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988
 A:Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ
 A:Reference number: A30796; MUID:88309090; PMID:2457365

A:Accession: A30796
 A:Molecule type: mRNA
 A:Residues: 266-421, 'V', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', '7
 A:Cross-references: GB:M21964; NID:g205685; PIDN:AAA41695.1; PID:g205686
 R:Leberburg, I.; Spliner, N.; Snyder, S.; Anderson, J.; Goldhaber, D.; Smulowitz, M.
 Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989
 A:Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament pepti
 A:Accession: A32757; MUID:89184647; PMID:2928342

A:Reference number: A32757
 A:Accession: A32757
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIE>
 A:Cross-references: GB:J04517; NID:g205679; PIDN:AAA41692.1; PID:g205680
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.1%; Score 197.5; DB 1; Length 1072;
 Best Local Similarity 22.1%; Pred. No. 0.03;
 Matches 112; Conservative 82; Mismatches 235; Indels 77; Gaps 20;

QY 11 AAAPGNGPEORLDYERAAIIGGPEDEGAFAHLPNRRK---LKEPGPLASSGGS 66
 DB 548 AEAKSPABAKSPAEEKSPATYKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPV 606
 QY 67 PAPSPAGCGGGGRLLPAGA-----APGQOESMGSGVPLPCPPATKQAGIGGE-P 118
 DB 607 PAEVKSPAEAS-----PAEAKSPAESVSPGFAKSPAEEKSPAEEKSPAEEKSPV 661
 QY 119 AAAGAGCSPRKYQAVLP-IGSGLVAAKEPTPMAGDKGAASPAATASDPAAPPPLPLP 178
 DB 662 AEVKSPTVYKSPAEEKSPVYKVS-PAVKSPE- AKSPAEEKSPAEEKSPAEEKSPA 719
 QY 179 GPPPLPATATAGTLAASGRKMSKSPVLGGGGGSGAASQAACLKQILLDLLEQOQO 238
 DB 720 SPAEAKPPAEAKSPAEA-----KSPAEEKSPAEEKSPAEEKSPAEEKSPV---EVKSPEAKS 768

```

Oy      239  QLOAKEIEIELSEEDOTLLARIERNERMOVLVKDKNEKHNKLFQGETEEREETLSE 298
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      769  PVEEGAKSLAEAKSPKCA-----KSPYKEELTPRAVEKSPERAKKSPME 812
Oy      299  KILCEOPELSETSOTLLPPKPPSCGSHGKHKRSFPGSTER-KTPYKLLAPEFSKYTK 357
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      813  EAK---SPEAKTALLDVKSPEAKTPAK--EAKRAPDISPEQVKSPAKEEKSPEKETR 867
Oy      358  TPKHSPKPEPCGSLSETVCK---BELNSOETPEKPRSSV-----DPPRLTSP---OK 405
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      868  TEKVAKPKBEVKS PVEEVAKEPRPKVVEEKTPATPKTEVEESKDKDEPKPAQPKFAEEK 927
Oy      406  GPSTH-PKEKAKSSEIEDLPLYLSTTEMYLCRHOPRPSPLRL-ESSPKKEETVARCIMP 463
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      928  EPLTEPKPDSPGEAKKEE-----AKKKAAAPREETPAKLVKEBAKPKKAAEDAKAKER 982
Oy      464  SVA-----GETSLVPSMRDHSYE 484
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      983  SKPSEKPKPKKEEVPAPAKKDTKEE 1008

```

RESULT 6

Bassoon protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: Y12730
 R:Dieck, S.; Sammetti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex
 J. Cell Biol. 142, 499-509, 1998
 A:Title: Bassoon, a novel zinc-finger CAG/Glutathione-repeat protein selectively localized
 A:Reference number: Z22249; MUID:98345363; PMID:9679147
 A:Accession: Y12730
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3942 <DIE>
 A:Cross-references: EMBL:Y17034; NID:93413809; PIDN:CAA76598.1; PID:93413810
 A:Experimental source: strain 129 SVJ
 C:Genetics:
 A:Map position: 9F1
 A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
 A:Note: bassoon
 C:Function:
 A:Description: may be involved in cytomatrix organization at the site of neurotransmitt
 A:Note: component of the presynaptic cytoskeleton
 C:Keywords: coiled coil; zinc finger

Query Match	6.1%;	Score 197.5;	DB 2;	Length 3942;
Best Local Similarity	22.5%;	Pred. No. 0.12;		
Matches 138;	Conservative 62;	Mismatches 173;	Indels 241;	Gaps 32

QY	32	GGPDEDEPAAVHFLPRHRKLTKEBGPCLASSOG-----GSPAP-----	69
Db	2125	GGPD-----LVQYOPQH-----GPGLSAQGLAPLRSGILGNPTYPEGSPSPGNLAQY	2172
QY	70	SPACGCGKGRGLLPA-----GAAPGQ-----OEE-----	95
Db	2173	GPAASQARNAVQQLLPSTATVAAAGDMYSTINTPLAATLPTTQPPASLRPMVKGCMTRP	2232
QY	96	--WGG--SVPL-----PCP-----PPATKQAGIGGEPA--	119
Db	2233	YVSGGVTAVPVLTLTRVPMIAPRVPLGAPGLYRPAPAPFPYASSVPAPEGVYL--GKRAA	2291
QY	120	--AAGAGCSPRPKYQAVLPYIQTGSLVAAKEPTPWADCKGAASPAATASDPAGPPPLPL	177
Db	2292	TKASGAGGPPRPPLPAGV-----AREPFF-----STTAPAVIKCAPVAPAPCPA	2335
QY	178	PGPPPLPATATAGTLASEGRWKKMKRSPDGGGSGGSSQAACIKQILLDLIDLEQQ	237
Db	2336	PAPPP-----GCKPAGEA-----AAGSGSGVLSRPAASEKE-----EASQ	2369
QY	238	QQQAAKEKEIEELKSERDPTLLARTERMERRRQYKKDKKEKRHLPLFCGYETEEKEEPTLS	297
Db	2370	EDRRKQOE--QLLQIERE---PVELEKLRQLRIQEELEERVEL-----QRHNEEDLL	2419

[illegible]

RESULT

tegument protein 24 - equine herpesvirus 4 (strain NS80567)
 C:Species: equine herpesvirus 4
 A:Variety: strain NS80567
 C:date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42567
 R:Telord, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 J. gen. Virol. 75, 1197-1203, 1998
 A:Title: The DNA sequence of equine herpesvirus-4.
 A:Reference number: 722173; MOID:98264497; PMID:9603335
 A:Accession: T42567
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3534 <TEL>
 A:Cross-references: EMBL:AF030027; NID:92605950; PIDN:AMC9539.1; PID:92605967
 A:Experimental source: strain NS80567
 C:Genetics:
 A>Note: 24
 C:Superfamily: varicella-zoster virus gene 22 protein

Query Match	5.98;	Score 192;	DB 2;	Length 3534;
Best Local Similarity	20.5%;	Pred. No. 0.19;		
Matches 158;	Conservative 88;	Mismatches 271;	Indels 252;	Gaps 36

```

OY      5  SAVKAAAPAGNGEOLDYERAAALGPDDEFGAANHLPRHRKLTKEGP-----P 58
Db      2717 TABSPKPAAPAPSKPAAPAPSKPAAPAP--SKPAAPAPSKP-----AAAPAPSKPAAP 2771
OY      59  LASSOGGSPAP-S-PAGCGGKRGILLPAGAAPGQOEESGGSVPLCPPPATKQIGGE 117
Db      2772 APSKPAAPAPSKPAAPAPSKPAAPAPSKP-----AAAPAPSKPA--APAPSK 2820
OY      118  PAAAGA-----GCSPPKXYAVLPJQGTGLVAAKEPTWADDKGAASPA-----ATASDP 169
Db      2821 PAAAPAPSKPAAPAPSKPAAP--APSKPAAPAPSK-----PAAAPAPSKPAAPAP 2874
OY      170  AGPPLPLPBPPIAPATACTLAASEGRMKSMRKSPJGGGSGSQAACLKOLLLO 229
Db      2873 SKPAAPAPSKPAAPAPSKPAAPAPSK-----PAAAPAPSKPQNTLVA 2917
OY      230  LDLIBQOOOOL--QAKKEIELELSEROTLLARIEMERRQOLYK---KDNKEKHK--- 281
Db      2918 IVAKDQAKDQAKDQAKDQAKDQAKDQ-----AKQAKQAKDQAKDQAKDQAKDQAKD 2974
OY      282  -----LRQGYETEE-----RETELSEKTKLSCOPELSETOTLPKPPESC 322
Db      2973 LTKOKSNAPAKFTGETPTLPNTSSGANPEMTPLDDPPIAVBP-----NPLPDD 302
OY      323  GRSKGHRKRSFPGSTERKTPVKKLAPBESKVTKTP--KHSPIKEBPGSLSETVQCR 379

```

```

Db      3025 -----DSPIGAVENTPLPDDSPIGAVENTPLPDDSPGLSPHQPVSRTLHNT 3075
QY      380 ELRSOE-----TPKPR-----SSVTPP----- 398
Db      3076 NIVSSDRSAAANVLPDPSDGFYSYAVNIPDPSPDDEFFSNOSRAQASAGSVSGS 3135
QY      399 -----RLSTPOKGPSTH-----PKKAFSSEIEDLPY-----LSTTEMYLIC-----R 435
Db      3136 SYKINTGRIPTAQAQFAHSHGRSRNRSTSKPSQSAFVKVPALSTKIPAPVPMOS 3195
QY      436 WHQPP-----SPDLRESSPKKEETVAR----- 459
Db      3196 HHAKGPNKPKCDGPTVLFSGSRNISPSQSTSTANISSTLPONOSTAKSHKVAKTPL 3255
QY      460 CLMPSSV-AGENTSVL-----AVPSMRDHSVEPLRDPNP-----SPLLENL-----D 499
Db      3236 RVPSSMPADDIDELPYDLQINRAVSNTPKPLPKSPQOPQPEYSSVTYDKONVRPMSE 3315
QY      500 DSVFSKRRAKLELDEKRRKMDIORIREQRIILQRLQRLMYKKKGIOESEPEVTSFPEPD 559
Db      3316 DEIALLINMNDTENDAEPIIDIKSIRQNLPKQI-----KQANKFVP--LDWWTETE 3367
QY      560 ---DVESLMTPEPLPVAVAFGRPLPKLTPQNFELPWLDESKRCRLEIQKK 605
Db      3368 SAADADGELSP-----KQPKL-----FSWESKRDLNINMLKDK 3401

```

RESULT 8

OPMSH

neurofilament triplet H' protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Jun-1999

C:Accession: J00368: A43778: S42616

R:Julien, J.P.; Cole, F.; Beaudet, L.; Sidky, M.; Flavell, D.; Grosfeld, F.; Mushynski,

Gene 68, 307-314, 1988

A:Title: Sequence and structure of the mouse gene coding for the largest neurofilament

A:Reference number: J00368; MUID:89121513; PMID:3220257

A:Accession: J00368

A:Molecule type: DNA

A:Residues: 1-1087 <JUL>

A:Cross-references: GB:M23349; GB:M24496; NID:g200034; PIDN:AAA39813.1; PID:g387493

R:Sheldman, P.S.; Carden, M.J.; Lees, J.F.; Lazzarini, R.A.

Brain Res. Mol. Brain Res. 4, 217-231, 1988

A:Title: The structure of the largest murine neurofilament protein (NF-H) as revealed by

A:Reference number: A43778

A:Accession: A43778

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 'M', 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545, 'R',

A:Cross-references: EMBL:J31012; NID:g463249; PIDN:CAA83229.1; PID:g463250

C:Genetics:

A:Gene: nfh

A:Introns: 290/1; 356/3; 398/2

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil; cytoskeleton; heterotrimer; intermediate filament; nerve; phosph

F:1-97/Domain: amino-terminal <NTE>

F:98-408/Domain: rod #status predicted <ROD>

F:409-1087/Domain: carboxyl-terminal <CTE>

F:519-886/Region: 6-residue repeats

F:520,526,532,538,544,550,556,562,568,574,580,586,592,598,604,610,616,622,628,634,640,64

73,685/Binding site: phosphate (Ser) (covalent) #status predicted

F:772/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match

Best Local Similarity 5.9%; Score 190; DB 1; Length 1087;

Matches 125; Conservative 52; Mismatches 222; Indels 136; Gaps 25;

```

QY      11 AAPAGNPEQLDYERAAALGPEDEPCAANHLPRRK-----LKEGPPLASGSGS 66
Db      636 ATVKSPEAKSPSEAKSPSEAKSPSEAKSPSEAKSPSEAKSPSEAKSPSEAKSPSEAKSP 694
QY      67 P-----APSPACGGCGRGILLPAGAPCOEESWGSVPLPCPPATKQAGIGCEP 118
Db      695 PAEVKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 743
QY      119 AAAGAGCSPRPYQAVLPQTGSLVAAAKEPTPMAGDKGAASPAATASDPAPPLPLP 178
Db      744 GEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 797
QY      179 GEPPLAPPTACTLAASBEKSKSMKSPGLGGGSGSAGAACLQIILLQDLIEQQQ 238
Db      798 IKPP-----AEAKSPSEAKSPYKEGAKP-----PEAK 825
QY      239 OQAKETEIEELKSERDTLLARIERMERQLVKKDNEKRLKFGYTEEREETLSE 298
Db      826 PLDVKSPEAQTVQEAATVPTDIRPEQ-----VKSPEAKKAK-----SPKEAKTSE 874
QY      299 KI--KLECG-BELSETSQTLPPKPFSCGSGSGKHKRSFPGSTERKTPVKKLAP-EPK 353
Db      875 KVAPKKEEVKSPYKEEVKAKPEPKVE-----EKTLPPTKTEAKSKKDEAPKAPK 927
QY      354 VKTTPKHSPIKEEPCGSLSEVYCKRELRSQETPEKPR---SSVTPPPLSTPOKGPSTH 410
Db      928 PVEERKETP-TEKPDSTAEK-----KKEAGKKKAVAAEETPALGVKEE---AK 977
QY      411 PEK--AFSSEIEDLPYLSTTEMYLCRWHPQPPSPPLRESPPKKEETVARCLMPSSVAG 468
Db      978 PKEKETTTTEADT-----KAKPSK-PTTEKPKKE-----MP----- 1012
QY      469 ETSVLAVPSWRDHSVEPL--RDPNPSDLE--NLDVSFSKRRAKLEDEKR 517
Db      1013 -----AAPKKDKTEKTESRKPPEKPKMEAKVKEDKSLSEKSPKTEAKK 1062

```

RESULT 9

T08599

probable transcription factor CA150 - human

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 18-Aug-2000

C:Accession: T08599

R:Sun, C.; Hayashi, T.; Liu, Y.; Lane, W.S.; Young, R.A.; Garcia-Blanco, M.A.

Mol. Cell. Biol. 17, 6029-6039, 1997

A:Title: CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, 1

A:Reference number: Z16449; MUID:97459702; PMID:9315662

A:Accession: T08599

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1098 <SUN>

A:Cross-references: EMBL:AF017789; NID:g2460123; PIDN:AAB80727.1; PID:g2460124

A:Experimental source: cell line HeLa

C:Genetics:

A:Gene: CA150

C:Function:

A:Description: HIV-1 Tat transcriptional coactivator

C:Superfamily: WW repeat homology

C:Keywords: nucleus; transcription regulation

F:131-168/Domain: WW repeat homology <WW1>

F:429-466/Domain: WW repeat homology <WW2>

F:527-565/Domain: WW repeat homology <WW3>

Query Match

Best Local Similarity 5.9%; Score 190; DB 2; Length 1098;

Matches 133; Conservative 80; Mismatches 221; Indels 238; Gaps 30;

18 NP-EQRLDYERAAALGPEDEPCAANHLPRRKLEKGPPLASGSGSPAPSPAGCG 76

15 NPGELRMAQQALRERGPAPPPNNAVVRGPP---LMPRPPEFGMRGPPPPRP----- 65


```

QY      175  -----LPGLPPLPATATAGTATLASEGKXKMSRSPJG--GGSGASASSOAAKQIIL  227
Db      364  SPTGSPSPPTGSPPTGPPRPGGKXSESSSE--RGPRGPRGPRGPR-----411
QY      228  LQDLLEQ000LOAKKEIELEKSERDTLLARIEMRMOVLKDKNEKRHLFOGE  287
Db      412  -----RKXSEE-----S  418
QY      288  TEREETELSEKIKLECOPELSETSQLPRPKFSGSGRGKHKRKKSPGSEKRTPYKKL  347
Db      419  SESEKEPR-----GPRSPPTGSPPTGSPPTGPRPG--SPVSGSPGLP----462
QY      348  APEFSVKTKTRHNSPIKEEPCSGSLSETVCKR---ELRSQETPEKPRSSVDTPRLSTP  403
Db      463  ----SROKQAPEDRPTGSPPTGSPPTGRPHNGDGGKXSESESEDEGPRGRSPPTGSP  518
QY      404  QKGPST-HPKKKAFFSEIED  422
Db      519  TGSPTGAPPKKGFASFLD  538

```

RESULT 12

eyleid - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: J13049
R:Freisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
Submitted to the EMBL Data Library, March 1998
A:Reference number: 217592
A:Accession: J13049
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TREE>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
:Keywords: DNA binding

Query Match	5.8%	Score 188;	DB 2;	Length 2715;
Best Local Similarity	21.4%	Pred. No. 0.22;		
Matches 125; Conservative	41;	Mismatches 175;	Indels 242;	Gaps 25;

```

0Y 11 AARPAGNENQOLDYERAAALGPRDECEGAANHLPRNKTKK--PGSPPLASSO--GGSP 67
Db 70 APPRPSGR-----GGRRPDRPAVMNHNLHQOOQONPPRRHHMOQHNGR 117
0Y 68 APSRACGGKGKGLLPAGA---ARGOEESMGSVLPCRPRA-----TK 110
Db 118 APPR-----PGSAPRNAHGVKEEY--TILPRPHNPAYGRYNADPNMDPYRG 163
0Y 111 QAGIGE-----PAAGAGSP--RKXQAVLPIQ-----IGSYVAAKE 148
Db 164 QIPRGKPRPOQOOPHEQOOPROCPRGSGSNPRPOQRITICQPOROGPTPTINSLQSSNP 223
0Y 149 PTF-----WAGD-----KCGASPARATASDPRAPRLPLPRPPLATPATGTLAEGR 198
Db 224 PRPRGNRYANTYPDQQAASAAMAAAAAQOQGAPRRGHNRPP--PDHNSPVGGGGG 281
0Y 199 W-----KSMKSPYL-----GGGGSGCASSOACLKOLLILLQL 230
Db 282 WABPRPRYSPOLPISOQYTPRTPTNTSRQOSYPRAHONSQSYSPSE-----329
0Y 231 DLLEQOONOLOAKEREIEELKSERDTLLARIEMRMQVLKKNEKERHKLFGYTEEE 290
Db 330 ---QQOQOQOQOQOQOQDA-----GQOQGGVPGGPPGT 355
0Y 291 REETELSEKIKLECOPELTSOTLP----PKPSCGRSGKGNK-----RKS 333

```

```

Dh 360 GQQR-----QONPRPTQSYRYRQRYPTRGRLAGGSNNRTAYSTHQYERRRW 410

Qy 334 PFGSTERKTVKKLADFSKTKTKPKH-SPIKEPQGSLEYVKRELNSQETPEKPRS 392
   |||
Db 411 PGSSSPSGGSHLP-----ASPHNVPLOOP-----PRPPHV 445

Qy 393 SVPTPRRLSTPQGRPTNRKEKAFSELEDRYLSTTEMYLCKRNORRPSRLRESRK 452
   |||
Db 446 SAGGP-----PSSSP-----GHASPSPQSPQSPSP 473

Qy 453 KEETVARCLMPSSVAGETSVLAVPSWRDHSVLEPLRNPDDL 495
   : : : ||| : ||| :
Db 474 HOELIGQNSNDSSSGAHSGM-----GSPFPQTPRPOQYM 508

```

RESULT 13

Hypochemical Caeprotein C18H7.3 - Caenorhabditis elegans
 C.Species: Caenorhabditis elegans
 C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C.Accession: T33110
 R.Ritt-Wollam, A.; Fronick, W.
 submitted to the EMBL Data Library, May 1998
 A.Description: The sequence of C. elegans cosmid C18H7.
 A.Reference number: Z21284
 A.Accession: T33110
 A.Status: preliminary; translated from GB/EMBL/DBD
 A.Molecule type: DNA
 A.Residues: 1-60 <TIN>
 A.Cross-references: EMBL:AF067607; PIDN:AAIC17641.1; GSPDB:GND0022; CESP:C18H7.3
 A.Experimental source: strain Bristol N2; clone C18H7
 C.Genetics:
 A:Gene: CESP:C18H7.3
 A:Map position: 4
 A:Introns: 84/1
 C:Superfamily: Phaseolus glycyne-rich cell wall protein 1.8

Query Match	5.8;	Score 186.5;	DB 2;	Length 460;
Best Local Similarity	29.0%;	Pred. No. 0.041;		
Matches 81; Conservative	14;	Mismatches 111;	Indels 73;	Gaps 12;

[illegible]

RESULT 14

hypothetical protein B7F21.40 [imported] - *Neurospora crassa*
 C.Species: *Neurospora crassa*
 C.Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 C.Accession: T51023
 R.Schulze, U., Aign, V., Hohnleisel, J., Brandt, P., Fartmann, B., Holland, R., Nyakatu
 submitted to the Protein Sequence Database, July 2000
 A.Reference number: Z5286
 A.Accession: T51023
 A.Status: preliminary

A:Molecule type: DNA
 A:Residues: 1:2649 <SCH>
 A:Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
 A:Experimental source: BAC clone B7F21; strain OR7A4
 A:Genetics:
 A:Gene: NCSP:B7F21.40
 A:Map position: 6
 A:Introns: 1619/3; 2584/1

Query Match 5.7%; Score 184.5; DB 2; Length 2649;
 Best Local Similarity 23.2%; Pred. No. 0.32;
 Matches 118; Conservative 60; Mismatches 171; Indels 159; Gaps 28;

```

QY      34 PEDPFGAAEAHFLPR---HRKL-----KEGEPPLASQGGSPAPSPAGCGGKGLPLPA 85
Db      1656 PTSPAGPRKRYDVPSSGHRPLAAAEAEFPDP--AKSEAAAPPAQPP-----SRFYQTPI 1708

QY      86 GAAPGQDESMGSGVPLPCP--PRATKQAGIGGEPAAAGAGCGSPRPXQAVLPITGSLVA 144
Db      1709 QASPVLTHTLVORSEFTVPAPLPRA-----PVAASASAPSPQ---TPSRPGAVS 1755

QY      145 AAKEPTWAGDKGGAASPAAT--ASDPAGPPPLPLPGPP-----PLAPATAG 190
Db      1756 QTMSPVPHP-----LRQPTATFVTEREG--EPVSGPPAQTQHPYRISQKTPAPVSSS 1809

QY      191 TLAASEGRKMSMKRSPGSGGSGASQAACLKQILLDLIEQQQQLQAKKEITIEL 250
Db      1810 MPASSEA-----MPRSAGMIPKANRPPLLSQ-----QHLEFREVEI 1846

QY      251 KSEEDTLARIERMERMOYVKKDKNEKRRHKLFGQYETERETELSEKIKLEQPELSE 310
Db      1847 REGEROI--RHEPLQPREQPPMERPRAMRK-----QEPE-----QPLHHD 1886

QY      311 TSQT--LPPKPFSGSGSGKHKRSPFSGSTERKTPVKKLAEFSKVKTKTPKHSPIKEPC 369
Db      1887 PFGWMPHPQ-----RSMQPRAAEAAPK---RQDEPRMAP--SAPQYTP--PIQAPV 1933

QY      370 GS--LSEYCKRELRSQCTPEPKPRSSVDTPPRLST-----PQ 404
Db      1934 RNLLSESV-----PPORTPOLSPAMERPVPSPQRPMPSTMOEQYSANTSAAQVPVPQ 1988

QY      405 KGPSTHPKEKAFSEIEDLYLSTEMYLCRMHQPPSPPLRLRSSPKKEETVARCLMPS 464
Db      1989 APAPIPEPKKVSISFILN-----DDNPAPAP---AKRVNDVAS--MPR 2030

QY      465 SVAGETSVLAVPSWRHSHVEPLRDPNPS 492
Db      2031 AASTST-----PPQQKSARPPQPPPT 2053

```

RESULT 15

T00273
 hypothetical protein KIAA0595 - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00273
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
 DNA Res. 5, 31-39, 1998
 A:title: Prediction of the coding sequences of unidentified human genes. IX. The complete
 A:Reference number: 214086; MUID:8290545; PMID:9628581
 A:Accession: T00273
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1520 <NAG>
 A:Cross-references: EMBL:AB011167; NID:g3043713; PIDN:BAA25521.1; PID:g3043714
 A:Experimental source: brain
 C:Genetics:
 A:Note: KIAA0595

Query Match 5.7%; Score 183.5; DB 2; Length 1520;
 Best Local Similarity 20.6%; Pred. No. 0.2;
 Matches 153; Conservative 64; Mismatches 216; Indels 311; Gaps 34;

```

QY      13 APAGNPEORLIDERRAALGPEDEPGAEAHFLPRRRKLEKEGPPPLASSQGGSPAPS-- 70
Db      681 SPASPEPEPVSKPVASS---PTEQVPSQEMPLLAR-----PSPVQSVPAVPPTPSM 731

QY      71 ----PACGGKGRGLLPAGAAPQDESMGSGVPLPCPPTATQAGIGEP-----AAA 121
Db      732 SALPFPAGGLGMPSPSLP---PPLOPPSLPLSGVPLPDFTHTYALPLPSMPCPHVSPS 788

QY      122 GAGCSPRPXYQAVLPITGSLVAIAAKERT--PWAGDKGGAASPAATASDPAGPPPLPLP 178
Db      789 GYCELPFP---PVPPLVSGPPGATVAPPTGSPVA-----PPAPV- 826

QY      179 GPPPLATATAGTIAASEGRKMSMKRSPGSGGSGASQAACLKQILLDLIEQQQ 228
Db      827 --SPYSSTCTYG-----PLGWPQ----- 843

QY      239 QLAKEKEIEELKSERDTLLARIERMRMOYVKKDKNEKRRHKLFGQYETERETELSE 298
Db      844 ----- 843

QY      299 KIKLECOPELSETSLPPKPF--SCGRS---GKHKRSKPSGSTERKTPVKKLAP-- 349
Db      844 ----PQNAPEWSTVPPRPPLPASIGRAVPQPKMSKGTGPRAPRENVLPISMARPLSL 896

QY      350 -----EFSKVKTKT-----PKH-----SP--IKEPQGS-----LSEYCKR- 379
Db      897 GLPCHGAPQTEPTEVEVKVPASPHPKHKSALVQSPOMKALACVSAEGTVVEPASERL 956

QY      380 ----ELRSQETPEKPRSSVDTPPRLSTPKGPSTHPKE-----KA 415
Db      957 KPEQTEPRKPEPLPATKAVPPRQSTVTKPLDVHVARLKLKLSFLPTPTQSGEDVQA 1016

QY      416 FSSSEI---BDL--PYLSTEMYLCRMHQPSP--LPLRESS---PKKEETVARCLMP 463
Db      1017 FISELTGEASDLSLLEQFEKSEAKKCEPPRAADSLAVNGSGVDIPQEKRLDRIQAP 1076

QY      464 --SSVAGETSVLAVPS--WRDHSVEPL---RDP-----NPSDLLENLSDVSFSKR 506
Db      1077 ELANVAGLTTPAPPPHQLMKPLAAVSLAKAKSPKSTAGCTLKPEGVTE-----AKH 1129

QY      507 HAKLEIDE-----KRRKMDIORIREQRIQ 532
Db      1130 PAAYRLOEGVHGSRVAVGSGDDHDYCRSRTPPKMPALVTPVGSMMVKKRHQDITIKP 1189

QY      533 RLQRLMYKKKGIOSESEPEVTSFPEPDD--VESLMTTPFLPVVAFGRPLKLTPOHFELP 590
Db      1190 VLSL-----GPAAPPPCIAASREPLDHTTSSEQADPSAPCLA--PSSLSP----- 1234

QY      591 WLDERSKRLLEIOKKQTPHRTCKR 614
Db      1235 ---EASPCRDMDMTRTPEPSAKQ 1255

```

Search completed: July 14, 2003, 11:51:29
 Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:07 ; Search time 72 Seconds
(Without alignments)
1136.332 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTWRSVAVFKAAPAGGNPE.....RSRCRLRIQKQTPHRTCRK 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1446	44.7	281	21	AA858317
2	1277	39.4	269	21	AA842665
3	377.5	11.7	195	22	AAU30711
4	226.5	7.0	720	23	AAE20839
5	226.5	7.0	766	23	AAE20838
6	226.5	7.0	973	23	ABG64675
7	226.5	7.0	973	23	AAE20799
8	219.5	6.8	699	22	AAAM1452
9	213	6.6	1038	22	AB871916
10	209.5	6.5	777	23	ABP41839

11	206.5	6.4	905	18	AAW31186	Human p160 polypep
12	206.5	6.4	1135	18	AAW31185	Human p160 polypep
13	206	6.4	1023	23	AAU82954	Human homologue of
14	205.5	6.3	676	22	AAU39666	Human polypeptide
15	203.5	6.3	2432	21	AAV55565	Human homologue of
16	199	6.1	617	22	AB835445	Peptide #2951 enco
17	199	6.1	617	22	AB820884	Protein #2883 enco
18	199	6.1	617	22	AAU56269	Human brain expres
19	199	6.1	617	22	AAU64458	Peptide #2892 enco
20	199	6.1	617	22	AAU4187	Peptide #2869 enco
21	198	6.1	1026	22	AAU78825	Human protein SEQ
22	196.5	6.1	2724	22	ABG20119	Novel human diagno
23	194.5	6.0	740	22	AAU93401	Human polypeptide,
24	194.5	6.0	1112	22	ABG22413	Novel human diagno
25	193.5	6.0	520	22	AAU52318	Murine WASP protei
26	193.5	6.0	520	22	AAU52318	Murine WASP protei
27	193.5	6.0	1033	22	AAU79809	Human protein SEQ
28	193.5	6.0	2263	22	AAU79000	Human protein sequ
29	193	6.0	753	22	AAU94374	Murine WASP protei
30	192.5	5.9	520	22	AAU9336	Human polypeptide
31	192	5.9	3536	22	AAU93273	Human polypeptide
32	188.5	5.8	554	23	ABU4721	Human p22464 prote
33	188	5.8	1798	21	AAU51611	Human HSGT1 protei
34	188	5.8	2703	22	ABU63299	Drosophila melanog
35	185.5	5.7	2819	22	AAU35408	Human 07CG27 gene
36	183.5	5.7	1502	22	AAU39273	Human polypeptide
37	183.5	5.7	1565	22	AAU41059	Human polypeptide
38	183	5.7	891	22	AAU79426	Human protein SEQ
39	183	5.7	1072	23	ABU57023	Mouse ischemic co
40	182.5	5.6	2897	22	ABU58514	Drosophila melanog
41	181	5.6	1231	22	ABU59466	Drosophila melanog
42	180.5	5.6	1685	22	ABG22718	Novel human diagno
43	178.5	5.5	2087	22	AAU31576	Amino acid sequenc
44	177.5	5.5	1798	23	AAU16277	Human kinase PRIN-
45	175.5	5.4	1413	23	AAU21729	Human PRIN-24 prot

ALIGNMENTS

RESULT 1	
AA858317	
AA858317 standard; Protein: 281 AA.	
ID	AA858317
AC	AA858317
XX	
DT	14-MAR-2001 (first entry)
XX	
XX	Lung cancer associated polypeptide sequence SEQ ID 655.
XX	
OS	Human: lung cancer associated protein; neuroprotective; cytosolic;
XX	cardiac; immunomodulatory; muscular active; vulnery;
XX	gastrointestinal; nephrotoxic; antiinfective; gynecological;
XX	antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX	proliferative disorder; wound healing; infectious disease.
XX	
XX	Homo sapiens.
XX	
XX	W0200055180-A2.
XX	
XX	21-SEP-2000.
XX	
XX	08-MAR-2000; 2000WO-US05918.
XX	
XX	12-MAR-1999; 99US-0124270.
XX	
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	(ROSE-) ROSEN C A.
XX	
XX	Ruben SM;
XX	
XX	WPI: 2000-587514/55.
XX	
XX	N-PSDB; AAF18193.

XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 11: Page 1153-1154; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548.. lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive; and
 CC immunomodulatory; muscular active general; vulnery; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 XX Sequence 281 AA;
 SQ
 Query Match 44.7%; Score 1446; DB 21; Length 281;
 Best Local Similarity 100.0%; Pred. No. 3.8e-89;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 181 PPLAPATAGTTLAASBGRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQQQQL 240
 DB 1 PPLAPATAGTTLAASBGRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQQQQL 60
 QY 241 QAQKEIEELKSRDITLARIEMERRMQLVKKDNKEKRLKFGYETERETELSEKI 300
 DB 61 QAQKEIEELKSRDITLARIEMERRMQLVKKDNKEKRLKFGYETERETELSEKI 120
 QY 301 KLECOPELSETSTLPKPFSCGSGKGRKSPFSTERTKPVKKLAPFESVVKTKTPK 360
 DB 121 KLECOPELSETSTLPKPFSCGSGKGRKSPFSTERTKPVKKLAPFESVVKTKTPK 180
 QY 361 HSPKKEPSCGSLSETVCKRELNSQETPERKRSVDPPLRLSTPOKGPSTHPKKAASSET 420
 DB 181 HSPKKEPSCGSLSETVCKRELNSQETPERKRSVDPPLRLSTPOKGPSTHPKKAASSET 240
 QY 421 EDLPYISTEMTCRWHOPPPSPPLPRESSPKKEEYVA 458
 DB 241 EDLPYISTEMTCRWHOPPPSPPLPRESSPKKEEYVA 278
 RESULT 2
 AAB42665
 ID AAB42665 standard; Protein; 269 AA.
 XX
 AC AAB42665;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2429, polypeptide sequence SEQ ID NO:4858.
 XX
 XX Human: Open reading frame: ORFX; detection; cytostatic; hepatotropic;
 KM vulnery; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antididiabetic;
 KM hypotensive; dermatological; immunosuppressive; antineoplastic;
 KM antiviral; antibacterial; antifungal; antineoplastic; antihypertoid;
 KM antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;

KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimketa RA, Leach M;
 XX
 XX WPI: 2000-602362/57.
 DR N-P-SDB; AAC76874.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 11: Page 4046; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antididiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antineoplastic; antihypertoid;
 CC antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 269 AA;
 Query Match 39.4%; Score 1277; DB 21; Length 269;
 Best Local Similarity 94.3%; Pred. No. 7.7e-78;
 Matches 247; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 197 GRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQQQQLQAKKEIEELKSEROT 256
 DB 5 GQMEYEYEEPSRQWMLRGSSSQAAACLKQIILLQDLIEQQQQQLQAKKEIEELKSEROT 64
 QY 257 LLARIEMERRMQLVKKDNKEKRLKFGYETERETELSEKIKLECOPELSETSTLP 316
 DB 65 LLARIEMERRMQLVKKDNKEKRLKFGYETERETELSEKIKLECOPELSETSTLP 124
 QY 317 PRPFSGSGKGRKSPGSTERKTPVKKLAPFESVVKTKPKHSPKKEEYVA 376
 DB 125 PRPFSGSGKGRKSPGSTERKTPVKKLAPFESVVKTKPKHSPKKEEYVA 184
 QY 377 CKRELNSQETPERKRSVDPPLRLSTPOKGPSTHPKKAASSETEDLPYISTEMTCRWH 436

Db 185 CKRELRSQETPEKERRSSVDTPPLRSTPOKSGPSTHPKKEKAFSSIEDLPYLSTEMYLCRW 244
QY 437 HOPPSPLPLRESSPKKEETVA 458
|||||
Db 245 HOPPSPLPLRESSPKKEETVA 266

RESULT 3
AAU30711
ID AAU30711 standard; Protein: 195 AA.
XX
AC AAU30711;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1202.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX

OS Homo sapiens.
XX
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US08656.
XX
XX 18-APR-2000; 2000US-0552929.
XX
XX 26-JAN-2001; 2001US-0770160.
XX
XX (HXSE-) HXSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20: Page 334; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AA029510-AA03304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 195 AA;

Query Match 11.7%; Score 377.5; DB 22; Length 195;
Best Local Similarity 46.3%; Pred. No. 1.1e-11;
Matches 94; Conservative 14; Mismatches 72; Indels 23; Gaps 8;

QY 10 AAAAPAGGNPEORLDYERAAALGSPDEPGAAEAHFLPRHRKLK-EPGPIASSQSGSPA 68
|||||
Db 1 AAAAPAGGNPEORLDYERAAALGSPDEPGAAEAHFLPRHRKLK-EPGPIASSQSGSPA 60
|||||

QY 69 PSPAGC-GGKGRGLLPAGAAP--GGOESWGSGSVPLPCD--PPATKAGIGG-----E 117
||||| ||||| ||| |::| ||| | : |||

Db 61 PSPAGCGGKGGGLVTPGRCGPRAAGREL----AVRCPCPVRRPPSPKALGSLDPOE 116
QY 118 PAAAGAGCSPRRKQAVLPLOTGSLVAAAEPTPWADDKGAASPAATASPAAPPLPL 177
||||| | : ||||| | : | : |||||
Db 117 PAAA-----PGPSIRPVLPLOTGSPWRRPSLRSLRPVLGTGVRGPRPLPPDPAPGAPPL 171
||||| | : |||||

QY 178 PGPEPLAPATATAGTL--AASEGR 198
||| | | : |||
Db 172 GPHPSRPPPTGPMRRARADGR 194

RESULT 4
AAE20839
ID AAE20839 standard; Protein: 720 AA.
XX
AC AAE20839;
XX
XX 01-JUL-2002 (first entry)
XX
XX Human gene 7 encoded secreted protein fragment, SEQ ID NO:101.
DE
XX Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
KW immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
KW rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
KW seazary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
KW Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
KW cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
KW thrombosis; ocular disorder; corneal infection; wound healing; cardiac;
KW vascular; thrombolytic; cyostatic; nootropic.
XX
XX Homo sapiens.
XX
XX WO200218435-A1.
XX
XX 07-MAR-2002.
XX
XX 17-JAN-2001; 2001WO-US01567.
XX
XX 28-AUG-2000; 2000US-228084P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI Ni J;
XX
XX WPI: 2002-281060/32.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
PT cardio-/cerebrovascular disorders and multiple sclerosis -
XX
XX Disclosure; Page 18-19; 504pp; English.
XX
XX AAD33237-AA033280 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
CC infections, anaemia, rheumatoid arthritis and multiple sclerosis),
CC cancers and hyperproliferative disorders (e.g. melanoma, neoplasms of
CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
CC infections). Secreted proteins of the invention can also be used to

CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 720 AA;

Query Match 7.0%; Score 226.5; DB 23; Length 720;
 Best Local Similarity 23.1%; Pred. No. 7.5e-07;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYERAALGCEDEPGAAEHFLRHRKLEKPPPLASSGGS-----PASPAG 73
 DB 50 PKRELEIQ-AQIKGPDVPGREDGKEAPEADLDRPGGIAVPVGEAHRHEPPVPHDKV 108
 QY 74 CGSGKGLLP-----AGAPG-----QGEESWGSGVPLPCPPA 108
 DB 109 VVDEGQDREVEENKPPSRHAGKAPGVGOMAPLPDSEREKQEPGEGVGKRPQAOA 168
 QY 109 TKOAG-----IGEPAAAGA-----GCSPPRYQAVLPLOTGSLVAAKE 148
 DB 169 LEEAGGLPDPKQVPEADQPAVQPAKEDLPGDRGHRPP--QAVLSQONGLAVGGGE 226
 QY 149 PT---PWAGDKGAASPAATASDPAGPPLPL--PGPPLPATATAGTLAASGKRWKSMR 203
 DB 227 KAGGPPPCNAADTQOPADSDHGKPPPAKPAAGPPLP-----EPRE 273
 QY 204 KSPILGGGSGGSSQ-----AALCKOILLDLLEQQOQQLAKEKEIEEKSERDTL 257
 DB 274 QRVVEAGAGQQAQLEBAGRAMLDHVALVQ-IEGQVQ-----KRLDQOEKL 324
 QY 258 LARIEKERRMQLVKKDKNEKRNKLFQGETEREETEISEKIKLEC-QPELSETSQTLR 316
 DB 325 LAVIEQDHKEIHOOORDEDEDKPR-----QVEVHOEPGAAPVPGQEAPEKAREYVENLP 379
 QY 317 PRFSGSGSGKHKRSPGSTERKTPVKKLAFESVVKTKPKHS--PIKEBSCSLSE 374
 DB 380 PLPLD-----PVLAPG--RPAPSDL-----NORSLEHSGPVRGAPAG--- 418
 QY 375 TWCKRELRSQETPEKPRSSVDPRLSTPGKPSHPKE---KAFSSELEDLPLSTTEM 431
 DB 419 -----PRDGPDTERRAAGKLRGQKNAARFRACTYK 451
 QY 432 YLCRMHQPPSPPLARESSPKKE 454
 DB 452 ELPKGPEOVVPPDAREAGPEE 474

RESULT 5

ID AAE20838 standard; Protein; 766 AA.

AC AAE20838;

DT 01-JUL-2002 (first entry)

XX Human gene 7 encoded secreted protein fragment, SEQ ID NO:100.

KM Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
 KM immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
 KM rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasia;
 KM Sezary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
 KM Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
 KM cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
 KM thrombosis; ocular disorder; corneal infection; wound healing; cardiant;
 KM vascular; thrombolytic; cytostatic; nootropic.

OS Homo sapiens.
 XX
 XX WO200218435-A1.
 XX
 XX 07-MAR-2002.
 PD
 XX
 XX 17-JAN-2001; 2001MO-US01567.
 PF
 XX 28-AUG-2000; 2000US-228084P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Edner R, Duan DR, Shi Y, Choi GH, Fischella M;
 PI Ni J;
 XX
 XX WPI; 2002-281060/32.

PT Isolated nucleic acid molecule encoding a human secreted protein 1s
 PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
 PT cardio-/cerebrovascular disorders and multiple sclerosis -
 XX
 XX Disclosure; Page 18; 504pp; English.

XX AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
 CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
 CC infections, anaemia, rheumatoid arthritis and multiple sclerosis),
 CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
 CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
 CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
 CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
 CC infections). Secreted proteins of the invention can also be used to
 CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 766 AA;

Query Match 7.0%; Score 226.5; DB 23; Length 766;
 Best Local Similarity 23.1%; Pred. No. 8.1e-07;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYERAALGCEDEPGAAEHFLRHRKLEKPPPLASSGGS-----PASPAG 73
 DB 96 PKRELEIQ-AQIKGPDVPGREDGKEAPEADLDRPGGIAVPVGEAHRHEPPVPHDKV 154
 QY 74 CGSGKGLLP-----AGAPG-----QGEESWGSGVPLPCPPA 108
 DB 155 VVDEGQDREVEENKPPSRHAGKAPGVGOMAPLPDSEREKQEPGEGVGKRPQAOA 214
 QY 109 TKOAG-----IGEPAAAGA-----GCSPPRYQAVLPLOTGSLVAAKE 148
 DB 215 LEEAGGLPDPKQVPEADQPAVQPAKEDLPGDRGHRPP--QAVLSQONGLAVGGGE 272
 QY 149 PT---PWAGDKGAASPAATASDPAGPPLPL--PGPPLPATATAGTLAASGKRWKSMR 203

Db 273 KAKGPPPPNAGDTGCPAEDSDHGKRPPLPAEKPAFGPLP-----EPRE 319
 QY 204 KSPFGGGGSGASSQ-----AACLKQILLDLIQOQOOLAKKEIKELSERDTL 257
 Db 320 QRDVERAGGNQASQLEBAGRAEMLDHVALVQV-IKSDQVQO-----KRLDDQOEKL 370
 QY 258 LARIERERRMQLVKKDKNEKERHKLFGYETEREREETELSEKIKLEC-QPELSETSOTLP 316
 Db 371 LAVIEBQHKEIHQQRDEDEKRP-----QVEVHQEGGAVPKQGEAPBEKARTEVENLP 425
 QY 317 PKPESCGRSGKHKRSPFGSTERKTPVKLAPFSVKTKTPKHS--PIKEPDCGSLSE 374
 Db 426 PLPLD-----PVLRAAPG--RPAPSQDL-----NQSLHSGPVRGDPAG---- 464
 QY 375 TVCKRELRSQETPEKPPSSVDTPPRLSTPOKGPSTHPKE--KAFSSEIEDLPYLSTTEM 431
 Db 465 -----PPDGGPDTERRAAGKLRDQCKDAAPRAAGTVK 497
 QY 432 YLCRMHQPSPPLPRLRESSPKKE 454
 Db 498 ELKPGPEQVPVDPDAREAGPEE 520

RESULT 6
 ABG64675
 ID ABG64675 standard; Protein: 973 AA.
 XX
 AC ABG64675;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Human albumin fusion protein #1350.
 XX
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antileukemia; antineoplastic; antitumor;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN W020017137-A1.
 XX
 PD 18-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US11988.
 XX
 PR 12-APR-2000; 2000US-229358P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2000; 2000US-256931P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Haseltine WA;
 XX
 DR WPI; 2002-010886/01.
 XX
 PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -
 XX
 PS Claim 1; Page 1418-1421; 2102pp; English.
 XX
 CC The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and

CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
 CC fusion proteins of the invention.
 XX
 SQ Sequence 973 AA;
 Query Match 7.0%; Score 226.5; DB 23; Length 973;
 Best Local Similarity 23.1%; Pred. No. 1.1e-06;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYTERAALGPEDEPGAAHFLPRHRRKKEKPPPLASSGGS-----PASPAG 73
 Db 303 PKERELEEO-AQIKGPVDVGRGDEKAPAEQAQDRRGQIGAVPGEAHNHEPPVPBDKV 361
 QY 74 CGGKRGGLLP-----AGAAPG-----QGEESMGSGVPLPCPPA 108
 Db 362 VVDEGQDREVPENKPPSRNAGKAPGVQGMAPPLDSEKKEKPEQGEVGRKPGQAQA 421
 QY 109 TKQAG-----IGEPAAAGA-----GCSPPKXQAVLPITQSLVAAKE 148
 Db 422 LEEAGDLPEDPQKVPADGGPAPVQPAKEDLGPDRLGLHPR--QAVLSEQNGLAVGGE 479
 QY 149 PT---PWAGDKGGAASPAATASDPAGPPPLP--PGPPPLAPVTAAGTLAASGRKMSMR 203
 Db 480 KAKGPPPPNAGDTGCPAEDSDHGKRPPLPAEKPAFGPLP-----EPRE 526
 QY 204 KSPFGGGGSGASSQ-----AACLKQILLDLIQOQOOLAKKEIKELSERDTL 257
 Db 527 QRDVERAGGNQASQLEBAGRAEMLDHVALVQV-IKSDQVQO-----KRLDDQOEKL 577
 QY 258 LARIERERRMQLVKKDKNEKERHKLFGYETEREREETELSEKIKLEC-QPELSETSOTLP 316
 Db 578 LAVIEBQHKEIHQQRDEDEKRP-----QVEVHQEGGAVPKQGEAPBEKARTEVENLP 632
 QY 317 PKPESCGRSGKHKRSPFGSTERKTPVKLAPFSVKTKTPKHS--PIKEPDCGSLSE 374
 Db 633 PLPLD-----PVLRAAPG--RPAPSQDL-----NQSLHSGPVRGDPAG---- 671
 QY 375 TVCKRELRSQETPEKPPSSVDTPPRLSTPOKGPSTHPKE--KAFSSEIEDLPYLSTTEM 431
 Db 672 -----PPDGGPDTERRAAGKLRDQCKDAAPRAAGTVK 704

QY 432 YLCRMHQPSPPLPRLRESSPKKE 454
 Db 705 ELKPGPEQVPVDPDAREAGPEE 727

RESULT 7
 AAE20799
 ID AAE20799 standard; Protein: 973 AA.
 XX
 AC AAE20799;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Human gene 7 encoded secreted protein HMVBP38, SEQ ID NO:61.
 XX
 KW Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
 KW immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
 KW rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
 KW seazary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
 KW Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
 KW cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
 KW thrombosis; ocular disorder; corneal infection; wound healing; cardiast;
 KW vascular; thrombolytic; cytostatic; neurotropic.
 XX
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT Peptide 1..34
 FT /label= signal_peptide
 FT Protein 35..973
 FT /label= Mature_secreted_protein
 XX W0200218435-A1.
 XX
 XX 07-MAR-2002.
 XX
 XX 17-JAN-2001; 2001MO-US01567.
 XX
 XX 28-AUG-2000; 2000US-228084P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M,
 PI NI J;
 XX
 XX MPI; 2002-281060/32.
 DR N-PSDB; AAD33243.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
 PT cardio-/cerebrovascular disorders and multiple sclerosis -
 XX
 PS Claim 11; Page 451-454; 504pp; English.

CC AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
 CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
 CC infections, anemia, rheumatoid arthritis and multiple sclerosis),
 CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
 CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
 CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
 CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
 CC infections). Secreted proteins of the invention can also be used to
 CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents a human secreted protein of the invention.

XX Sequence 973 AA;

Query Match 7.0%; Score 226.5; DR 23; Length 973;

Best Local Similarity 23.1%; Pred. No. 1.1e-06; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYTERAALGPEDEGAEAHFLPRHRLKKEGPPLASSOGS-----PAPSPAG 73
 Db 303 PKRELEIQ AQIKGEYVDVGRBDGKEAPEAQLDRGGCIANVGEAHNHEPPVPHDKY 361
 QY 74 CGGKGGGLLP-----AGAPG-----QOESWGSVPLPCPPA 108
 Db 362 VVDEGQDREYVENKPPSRHAGKAPGVGOMAPLPDSEREKQPEEGGVGRPGQAQA 421
 QY 109 TKQAG-----IGGEPAAGA-----GCSPPKTYQANVPIOTGSLVAAKE 148

Db 422 LEEAGDLPEDPQKVPPEADGQPAVQPAKEDLGPDRGIHPRP--QAVLSEQONGLAVGGE 479
 QY 149 PT---PWAGKKGGAAPAAASPAPRPLPL--PPPPPLPATATGTLAASGRKKSAR 203
 Db 480 KAKGPPPGNAADTQGPAPEDSDHGKPPPAKPPAPGGLP-----EPRE 526
 QY 204 KSPDGGGGSGGSSAQ-----AACLKQILLDILLIQOQQOQQAKEKEIEELKSBDTL 257
 Db 527 QRDVERAGGQAASQLEENGRALMDHAVILQY-IEQQVQO-----KRLDDQDKL 577
 QY 258 LARIEMERRMQLVKDNEKERHKLFGYTEEREETEELSEKIKLDC-QPELSETSQLP 316
 Db 578 LAVIEQHKKEIHQORODEEDKPR-----QVEVHQEPGAAPVPGQAPAEKAREVENVLP 632
 QY 317 PKPFCGSRSGKHKRSFPGSTRKTPVKKLAPEFSKVTKTPKHS--PIKEPCCSLSE 374
 Db 633 PLPLD-----PVLAPGQ--RPASDOL-----NORSLSESGPYGRDPAQ---- 671
 QY 375 TVCKRELRSQETPEKPRSSVDTPTPLSTPQKGSPTHKE--KAFSEIEDLPLYLSTTEM 431
 Db 672 -----PPDGGPTTEPRAAGKLRDGGQKDAAAPRAAGTYK 704

QY 432 YLCRWHPSPPLPLRESSEPKKE 454

Db 705 ELPKGEQYVPPDPAREAGPPEE 727

RESULT 8

AA041452

ID AA041452 standard; Protein; 699 AA.

AC AA041452;

XX 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6383.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

XX Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;

XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX chemokine; thrombolytic; drug screening; arthritis; inflammation;

XX Leukemia.

OS Homo sapiens.

XX W0200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000MO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEO INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX MPI; 2001-442253/47.

XX N-PSDB; AAI60608.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

Qy	430	-----EWYLCRHHQPPSPPLFLRESSPKKEVYARCLAMPSSVAGETSVLAVPSWDH	48
Db	842	DOEHCHDNFFL-----EEA-----DELLADNPSELETKMDV	87
Qy	482	SVEPLRDPNPSDLLLENLDDSVFSKRRAKLEDEKRRKMDIQRIRQIRLQRLRMVK	54
Db	875	PVPPSSDKRIDTEL--LSDATFERRRHQXVKVDEVRKCRKADAYMEOQLRDLRRRRND	93
Qy	542	KGIOESP-EVTSFFPEPDDVSLMTTPPLPVVAGRPLPKLTPD-NRELPMLD	593
Db	932	EVVLADPLRASFTFPLPDLIEALIOFVNEVTVOAEGENVVMNEADDGFGVWVD	985
RESULT 10			
ID	ABP41839	standard; Protein; 777 AA.	
XX	ABP41839;		
AC			
XX			
DT	22-AUG-2002	(first entry)	
XX			
DE		Human ovarian antigen HNBV053, SEQ ID NO:2971.	
XX			
KW		Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;	
KW		ovarian cancer; breast cancer; tumour; reproductive system disorder;	
KW		infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;	
KW		PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;	
KW		inflammatory condition; immune disorder; blood disorder;	
KW		cardiovascular disorder; respiratory disorder; neurological disorder;	
KW		gastrointestinal disorder; urinary system disorder; drug screening;	
KW		gene therapy; chromosome mapping; forensic analysis;	
KW		antibody preparation; cytostatic; immunomodulatory; neuroprotective;	
KW		antiinflammatory; gynaecological; reproductive.	
XX			
OS		Homo sapiens.	
XX			
PN		MO20020677-A1.	
XX			
PD		03-JAN-2002.	
XX			
PF		07-JUN-2001; 2001MO-US18569.	
XX			
PR		07-JUN-2000; 2000US-209467P.	
XX			
PA		(HUMA-) HUMAN GENOME SCI INC.	
XX			
PI		Blirse CE, Rosen CA;	
XX			
DR		WPI: 2002-147878/19.	
XX			
DR		N-PSDB: AB054916.	
XX			
PT		Isolated nucleic acid molecules encoding novel ovarian polypeptides,	
XX		useful in the prevention, treatment and diagnosis of cancer (e.g.	
PT		ovarian cancer), immune disorders, cardiovascular disorders and	
XX		neurological diseases -	
XX			
CC		Claim 11: SEQ ID NO 2971; 2922bp; English.	
CC			
CC		The invention relates to 2175 novel human ovarian antigens (ABP41054-	
CC		ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also	
CC		encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC		to the sequences of the invention. The invention additionally relates to	
CC		recombinant vectors and host cells comprising human ovarian antigen	
CC		polynucleotides, antibodies against human ovarian antigens, and the use	
CC		of ovarian antigen polynucleotides and polypeptides in diagnosing,	
CC		treating, prognosing or preventing various ovary and/or breast-related	
CC		disorders. Such conditions include ovarian cancer and breast cancer, and	
CC		metastatic tumours of ovarian or breast origin, reproductive system	
CC		disorders (e.g., infertility, disorders of pregnancy, anovulation,	
CC		polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine	
CC		disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic	
CC		shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and	
CC		vaginitis), immune disorders (e.g. congenital and acquired	

Query Match	Best Local Similarity	Score	DB	Length
Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;	23.1%; Pred. No. 1.1e-05;	209.5;	23;	777;
QY	55	GPPLASSQGG-----	SPASPACCGCGKGLLLPAGABG--QOESWGGS	99
Db	347	GPPTTAHNLGVSGLVSPRLPGENHRAGSNEDPLASGTPPTTPDPTFGGR		406
QY	100	VPLECPRPATQ-----	AGGEBPAAAGACGSPRYQAVLPQ	138
Db	407	VPRAFYHYDEEASDVLSLESDDSVIVPEGLPRLPPPPGATPP-----		457
QY	139	TGSLVAAMAKKETPMAGDKGAASPAATASD-----	PAGPPPLPLGPPPLAPATATAGTLA	193
Db	458	-----IAPGPT-----	ASPPVPAKEEPEELPAAPGRLP-PPPPPPVPVCPVPLP	503
QY	194	ASEGKWKSMKSPGLGGGGSGASSQACOLKQILLQDLIDQOQOOLAKKEIEELKSE		253
Db	504	PPQ-----	LVPECTPGGGPPPALEEDLVITINISSD--EEEEEDEEEDEEEDEEE	554
QY	254	RDTLARIERNRKQOLYKKNKEKRNHLFGYETEEEBEETLSKITLCOPELSEFSQ		313
Db	555	ED-----	FEDEEEDDEEYFEDEEEDDEEFEDEEEDDEEELBEDEEEDDEEELBEV--	607
QY	314	TLPPKPFSCGSGSGHKRKSPPGSTERKTPYKTLAPERSKVTKTPKHSPIKEPCGSLS		373
Db	608	-----	EDLEFGTAG-----GEVEECAPPPPTPLPALPPESPSPKQOPEPEPPGLLL	654
QY	374	E-----	TVCKRELRSQ--ETPEKPRSSVDTPPRLSTPOKGPSTHPKERAF	416
Db	655	EEVEPGTEERKADPTAPLAPDEALPSQGEVEEBEGSPAAGRPPOGLVEEBSAPP--TL		712
QY	417	SSEIEDLPLSTETMYLTCRHHQPPSPPLAKSSKKKEETVARKCLMPSVAGETVLA		476
Db	713	EEETED-----	GSDDKV--QPPPE-TPAEEE-METETEAEALOEKODDOTAAMLA--	757
QY	477	SMRDHVSVEPLDPPNSD		493
Db	758	DFIDCPDDEKPPPE		774
RESULT 11				
AAW31186				
ID	AAW31186 standard; Protein; 905 AA.			
XX	AAW31186;			
XX	11-MAR-1998 (first entry)			
DE	Human p160 polypeptide 160.2.			
XX				
XX	p160; p62; cytoplasmic; T cell; B cell; development; activation;			
KW	modulation; cellular response; cell proliferation; autoimmune disease;			
KW	p56-lck.			
OS	Homo sapiens			

XX	Key	Location/Qualifiers
FH	Domain	3..138
FT		/label= leucine_zipper
FT	Region	510..638
FT		/note= "proline/lysine rich region"
FT	Region	654..870
FT		/note= "glutamic acid rich region"
PX		
PN		
XX	M09J22225-A1.	
PD	26-JUN-1997.	
XX		
PF	11-DEC-1996;	96WO-US19944.
XX		
PR	19-DEC-1995;	95US-O574959.
PA	(DAND) DANA FARBER CANCER INST INC.	
PI	Joung I, Shin J, Strominger JL, Vadmudi RK;	
DR	WPI: 1997-341351/31.	
N-PSDB:	AAT89346.	
PT	cDNA encoding p62 and p160 and corresponding proteins - used in the treatment of autoimmune disease and for T and B cell proliferation, e.g. for treatment of tumours	
PS	Claim 84; Fig 11; 175pp; English.	
XX		
CC	This sequence represents a novel p160 polypeptide (160.2) which is capable of activating transcription of a variety of genes upon activation of p62 and is capable of binding to the p62/p56lck complex to modulate lck function in a manner similar to p62. The genes transcribed include response to p160 activation likely include those of which are involved in T or B cell development/differentiation, T or B cell activation or production of T or B cell specific factors e.g. lymphokines or antibodies. This p160 polypeptide is also a substrate for serine/threonine kinase activity. p160 polypeptides can modulate degradation of cellular proteins e.g. cell cycle regulatory proteins stimulating expression of cell cycle dependent kinase inhibitors and arresting cell cycle progression at specific boundaries to thereby modulate cell proliferation. As p160 boosts B cell response it may be used to treat disorders where this is beneficial, e.g. infections by pathogenic microorganisms. p160 can be used to expand T cell populations for treating infectious diseases or cancer and p160 inhibitors could reduce B or T cell responses and may be used to treat a variety of autoimmune diseases, e.g. diabetes mellitus, arthritis, multiple sclerosis allergic reactions and Crohn's disease.	
CC		
SC		
XX	Sequence 905 AA;	
Query Match	6.4%; Score 206.5; DB 18; Length 905;	
Best Local Similarity	22.4%; Pred. No. 2.2e-05;	
Matches 121; Conservative	52; Mismatches 231; Indels 137; Gaps 21	
OY	38 PGAEAHNLPLPHRRK-----LKEEGPPLASSGGSPAPGCGKGRLLPAGAAG 90	
Dd	414 PTPAPVPLLRRHRHSRHSILRACPQMA-----PCDQAPCPSAG-----PMPSACP 462	
OY	91 QOESWG-----SVPLPCPP-----PATKAGIGGEPAAGACS----- 126	
Dd	463 VPSEFWSTYTANLGLSRPSVCPRRLPGPENRHAGSNEDTILAPSGRPPTIPDETF 522	
OY	127 -----PRPKY-----QAVLPITGS-----LYAAAKEPPLMAGDKGAASPAATASD 168	
Dd	523 GGRVPRPFAFVHYDKREASDVETLSDDSDSVIIVEGLPLPLPPPPSGATPEPIAATGP 582	
OY	169 PAGPPLPLU-----GPPPIATPATAGTLAASEGRMKSKRSPLGGGGS 213	
Dd	583 PTASPFPVAKEEPDELPAAPGPLPPPPPVPVLPPO-----LYPEGTPGGG 635	
OY	214 GASSOAACIKOILLQLDLIEGOOOOLAKKEIEELKSERDTLLARTERMERRMOVLVK 273	

Db	636	GPALLEDLTVININSDDSEEEEGEEEEE	EEEEEEEEEEEEEEEE	EEEEEDFEEEEE	6900
Oy	274	DNE---KERKILPQGYETE-EREETELSEKIKLE	CQPLSTETSTPLPKPPSCGRSGKH	329	
Db	691	DEEYFEEEEEEEEEFEFEFEFEFEFEFEFEFE	EEFEFEFEFEFEFEFEFEFEFEFEFEFE	EDLEFETAG---	7444
Oy	330	KRKSPEGSTERKTPVKKLAPESKVKTKTPKHSP	IKPECGSLSEYCKRLRSQETPEK	389	
Db	745	-----GEVEGAPPEPTLPPLAPPESP	PKVQPEPEPGILLEV-----	EEEGTE	790
Oy	390	PRSSVTPPLPLSTPQCGPSTHREKAFSSE	IEDLYLSTETMYLCRHHQPPSPPLURE-	448	
Db	791	EEEGADAPPLA-PEALPQGEVERGESPA	GPPPOELVE-----	EEESPPTLLLEE	843
Oy	449	-----SSPKE---ETVARCLMP	SVASGERSVLAVPMSRHSVEPLRDPNS	492	
Db	844	TEGSDKVPQPPETPAEEMETETEA	LAQEKEDDTAAMLA--DFIDCPPDDEKPPPT	901	
Oy	493	D 493			
Db	902	E 902			
RESULT 12					
AAW31185					
ID	AAW31185	standard; Protein; 1135 AA.			
AC	AAW31185;				
NC					
DT	11-MAR-1998	(first entry)			
XX					
DE	Human p160 polypeptide 160.1.				
XX					
KM	p160; p62; cytoplasmic; T cell; B cell; development; activation;				
KW	modulation; cellular response; cell proliferation; autoimmune disease;				
XX	p56-lck.				
OS	Homo sapiens.				
XX					
FT	Key	Location/Qualifiers			
FT	Domain	3..138			
FT	Region	/label= leucine_zipper			
FT	Region	740..868			
FT	Region	/note="proline/lysine rich region"			
FT	Region	884..1100			
FT		/note="glutamic acid rich region"			
XX					
PN	W09722255-A1.				
XX					
PD	26-JUN-1997.				
XX					
PF	11-DEC-1996;	96WO-US19944.			
XX					
PR	19-DEC-1995;	95US-0574959.			
XX					
PA	(DAND) DANA FARBER CANCER INST INC.				
XX					
PI	Young I, Shin J, Strominger JL, Vadamudi RK;				
XX					
DR	WPI; 1997-341351/31.				
XX					
DR	N-PSDB; NAT89345.				
XX					
PT	cDNA encoding p62 and p160 and corresponding proteins - used in the				
PT	treatment of autoimmune disease and for T and B cell proliferation,				
XX	e.g. for treatment of tumours				
PS	Claim 84; Fig 9; 175pp; English.				
XX					
CC	This sequence represents a novel p160 polypeptide (160.1) which is				
CC	capable of activating transcription of a variety of genes upon activation				
CC	of p62 and is capable of binding to the p62/p56lck complex to modulate				
CC	lck function in a manner similar to p62. The genes transcribed in				


```

OY 224 -----OILLQLDLIEQOQOOLAKK----- 245
DB 433 RLPOPPONTNIOFOLPFGVAVLYRSENGOULMIPQALMOQAOHAPOTTMAPRPAT 492
OY 246 -----EIELKSERDTLLAR-----IERMERMOVLKDKNEKER----- 279
DB 493 PTAPVQISTVQAAPRIIAROVTPPTIIKQVQAQOTTVQPSATLQSPGVQPOLVLGG 552
OY 280 -----HKLFQGYETEBEREETELSEKIKELCOPELSETSQTLPPKPF 321
DB 553 AAOGTASLGTAFAVOTGPORVPGATVTSNATEMENVK-KCKNFLS---TL---IK 603
OY 322 CGRSGKGHKRSPGSTERKTPVKKL-----APEFSKVTKPKHSPKIEEPGCS 371
DB 604 LASSGK-----QSTETAAVVELVONLDGKIEADFTSRLYRELNSSP---QP--- 649
OY 372 LSETVCKREL-----RSQETPEKPRSSVDPTRPRL---STFOKGPSTHPK 412
DB 650 YLVPEFLKRLPALRQLPFDNAFLQGSQOQPPPTSQATTLALVAVLSSVQR---TAGK 706
OY 413 EKASSSEEDLPYISTE-MYLCRWHPPPSPPLPRESPPKKEETVANCIMPSSVAGETS 471
DB 707 TAAVTVSALOPPVLSLTQPTGVGKOGQPPPLVIOQ--PPKPGALIR--PPOVTLIQTP 762
OY 472 VLAVPMSRHDHVEPLRDPNPSDLLENLDDSVFSKRHAKL-----ELDEKRRKMDIQIR 526
DB 763 MVA-----LRDPHNRIMLTTPQOVNLSSESARILATNSELVGTTRSCDETFEL 811
OY 527 EORILORLQRLMYRKKGIOESEPEVTSF 554
DB 812 LQAPLQRRRIETIGKKHGITELHPDVVSF 839

```

RESULT 14

AAM39666 standard; Protein: 676 AA.

```

XX AAM39666;
XX AC AAM39666;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 2811.
XX KW Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX XX 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HSEB-) HSEBQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

```

XX WPI: 2001-442253/47.
 DR N-PSDB: AA158822.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 4: SEQ ID NO 2811; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161365) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

SO Sequence 676 AA:

Query Match 6.3%; Score 205.5; DB 22; Length 676;
 Best Local Similarity 22.0%; Pred. No. 1.8e-05;
 Matches 131; Conservative 82; Mismatches 224; Indels 159; Gaps 25;

```

OY 55 PGPLASQGGSPAPRAGC---GKGGILLPAGAAQGOEEMSGSVPLPCPPRATKO 111
DB 2 PSAKORSGKGHGAASPEKGAHPSAARPLAAPPRAACSPSPGGA---PASPPGAP 58
OY 112 AGIGEPAPAAAGACSPRRKYQAVLPIOTGSLVAAKERTPPACDGG--GAASPAATADP 169
DB 59 RSLASQPARAAARAPM-----SAKQ---RSGSGHGAASPEKGAH 99
OY 170 AG-----PPPLPGLPPLA-----PRTAGTGLAASGRMKSMKSPSLGGGSGG 214
DB 100 SGGADVAAKPPRPAQOPPPRPAHPOOHPOHNGKGRH-----GGGGGGGK 152
OY 215 -----ASSQAACIQI-----LLQLDL-----IEQOQOOLAKE 244
DB 153 SSSSSASAAAAAASSASCRRGLALNFLYLVAALAAFGWCVHHVLEVOQVR 212
OY 245 KEIELKSERDTLLARIMERRMOLVKDKNEKERNKLFQGYETEEREETELSEKIKLEC 304
DB 213 RSHODFSKORELQGLGVQKQSL-----QATFGFESILSSQHKODLTAKVAKG 267
OY 305 QPELSETSQTLPP-----KPFSGSGGKHKRSPGCS-----TERKTVPYKLAPE--- 350
DB 268 ESEVSRISSEVLQKIONELIKDLSGDIHVKKDAREDFSLNTEVEERTELTKSINDNIA 327
OY 351 -FSKYTKTPKHSPTKEPCGSISFT-----VCKRELQSQETPEKPRSSVDPTRPLS 401
DB 328 IFTEVQKRSQKEINDMKAKVASLSESEGNKODLAKLKAKEIQTSAASR--EWDDEALRS 386
OY 402 TPQ-----KGPSTHPKEKAF--SEETEDLPYSTEMYLCKRNHPSPPLPRL 448
DB 387 TLQTMESDIYREVRILVSLKQEQAFKRAADTERLALALTEKLL----- 431
OY 449 SSPKKEETVANCIMPSSVAGETSVLAVPSMDH-----SVEPLDPPND--SDL 494
DB 432 ---RSESVSR--LPETIRLLEEELRQLKDSHGCKEDGGRHSEAFALOOKSGGLDSR 486
OY 495 LENLDDSVFSKRHAKLEIDKRRKMDIQIRREORILLORLQRLMYKKKGIOSEPE 550
DB 487 LQHVEDGLSMQVASAROTESLESILSKSQHEDR--LALQGRL---EGLGSSAD 538

```

RESULT 15
 ID AAY85565 standard; Protein; 2432 AA.
 XX AAY85565;
 XX 07-JUL-2000 (first entry)
 XX Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
 DE UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
 XX anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 XX anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT 1..89 /note="this region can be replaced with one of the
 FT three sequences shown in AAY85566 to AAY85568;
 FT this creates three variants at the N-terminus"
 FT Misc-difference 1018
 FT /label= Asp or Glu
 FT 1776..1778 /note="present or absent depending upon the allele
 FT from which the protein is translated"
 FT Region
 XX
 XX MO9963080-A1.
 XX
 XX 09-DEC-1999.
 XX
 XX 02-JUN-1999; 99WO-EP03848.
 XX
 XX 03-JUN-1998; 98GB-0011962.
 XX
 XX (JANC) JANSSEN PHARM NV.
 XX
 XX Luyten WHM, De Raeymaeker MC, Geysen JGH, Bogaert TAOE;
 PI Maerten LJS, Verhasselt P, Van De Craen M;
 XX
 XX WPI: 2000-116370/10.
 DR N-PSDB; AAA07836.
 XX
 XX Novel proteins and nucleic acids e.g. for treating neurodegeneration
 PT
 XX
 XX Claim 95; Fig 1d; 146pp; English.
 XX
 XX The invention provides vertebrate (human) protein homologue of a UNC-53
 CC protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
 CC their plus ends. The UNC-53 sequences are used to promote neural
 CC regeneration, revascularization and wound healing; also for treating
 CC neurodegenerative disease, acute traumatic injury, fibrotic disease and
 CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
 CC polynucleotides can be used for recombinant production of the proteins,
 CC as a source of probes for detecting allelic variants and polymorphisms,
 CC for sequencing genomic DNA and for detecting UNC-53 expression; and as
 CC source of therapeutic antisense sequences. Cells that express the
 CC protein are used to identify regulators of cell shape, growth, motility
 CC and migration. They can also be used to identify proteins that are
 CC involved in signal transduction pathways also involving UNC-53, and to
 CC identify compounds that alter attachment of UNC-53 to microtubules. A
 CC target gene coupled to a UNC-53 encoding sequence may be used to deliver
 CC the target gene to a cellular microtubule or its plus ends. The present
 CC sequence represents the amino acid sequence of the second human homologue
 CC of UNC-53, designated hs-UNC-53/2.
 XX
 XX Sequence 2432 AA;

Query Match 6.3%; Score 203.5; DB 21; Length 2432;
 Best Local Similarity 22.3%; Pred. No. 0.00012;
 Matches 130; Conservative 81; Mismatches 259; Indels 113; Gaps 21;

QY 6 AYKRAAARA---CGNPEQLDIERAAALGPEDEPGAAEAHFLPRHRKLKEPPPLASS 62

Db 209 AVSQVAGAPSCQAGTPOQOV-----PYTPQAPCCPHQAPHQSKAEMQSRL 238
 QY 63 QGSPAPAPAGCGCGKGRLLPAGAAPGOOESEWGSVPL--PCPPATQAGIGCPAA 120
 Db 259 SGPTARVSAAGSEKTKGTGGSTTANNRSQSEFNNDKSKPTYSPPPPSSHEK----EPLA 314
 QY 121 AGAGCSPPKRYQAVLPYTGSLVAAPAEPPPMAGDKGAA----SPAARAPDPAPPL 175
 Db 315 SSASSHPGMDNAPASLESGSSSTPTNCSTSAIPQGAATKPMRSKSLVKSATVSM 374
 QY 176 PLPAPPPLPATATGTLAASBGRKSMRK-----SPLGGGGSGASSQAACLKQILLIQ 229
 Db 375 SVKPPGEAPRPPEAKKPAANNQSMLEKTLKFNKSGSKAGSGPESRDTSCERETLP 434
 QY 230 LDLEQOQOOLQAKEKEIEEL---KSERDTLARIEMRMRLVKD---NEKERHK 281
 Db 435 ---SFESELELAASRLITVGPASSSPKIALKGIORTSRALTNRKSLKGRKEK 491
 QY 282 LFGYTEEREE---TELSEKIKLECPLESTSTQTLPPKPSGCRSG---KGHK-- 330
 Db 492 QOREKDKESKDLAKRASVTERLDLKEPKEDPSGAAVPMPKSKRIASFIPKGLNS 551
 QY 331 -RKSPGSTERKTPVKKLAPBFSVKTKTKK-HSPIAEPCGSLSEIVCKRELRSQPTPE 368
 Db 552 AKKEPMAPSHSGIP---KQGMKSMPSKSPAPAPSK-----GERSRSGKLSSGIPQ 601
 QY 389 KP-----SSVDTPLRLSTPOKGPSTHPKKEKAFSE-----IEDLPLSTEMVLCRW 437
 Db 602 KPOLDGHHSSSSSLASBEGKRGITLNNISISSQVSGVGTQTGGSTVAVOL----- 657
 QY 438 QPPSPPLPRESSPKKEETVARCIMPSSVAGETSVLAVPMSRDHVEPL----- 486
 Db 658 ----PQPOQVNHNPATVAPFLYRSQDTDEGNVTAESSSTGVSEPSHFTKTGPAL 712
 QY 487 ----RDP-----NPSDLENTLDSVSKR-----HAKLE 511
 Db 713 ELAGEDPEARRLRTVKNIAOLRONLETMSLSRGTOVTHSTLE 755

Search completed: July 14, 2003, 11:48:31
 Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 13:41:51 ; Search time 4488 Seconds

(without alignments)
3981.534 Million cell updates/sec

Title: US-10-054-935-2
3238
Perfect score: 1 MTRSAVFKAAPAGGNPE.....RSRCLEIQQKQPHRTCRK 614
Sequence:

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n-model -DEV-xlh
-O/cgn2_1/USPRO/spool/US10054935/runat_14072003_115025_24625/app.query.fasta.1.775
-DB-genemb1 -QPM-fastlap -SUFFIX-rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human4.0.cdi -LIST=45
-DOCLALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTWMT-plc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054935.ecgn.1.1.2819 @runat_14072003_115025_24625 -MCPU=6 -ICPU=3
-NO_MMAR -LARGESQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Genemb1:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1757.5	54.3	196724	10 AL590963	AL590963 Mouse DNA
2	1460	45.1	180017	2 AC068669	AC068669 Homo sapi
3	1311	40.5	193189	2 AC119462	AC119462 Rattus no
4	1055.5	32.6	615	6 AX401049	AX401049 Sequence
5	944.5	29.2	174346	2 AC131356	AC131356 Rattus no
6	906.5	28.0	193189	2 AC119462	AC119462 Rattus no
7	680.5	21.0	1433	9 HSM800235	AL049450 Homo sapi
8	416	12.8	207214	2 AC107643	AC107643 Mus muscu
9	380.5	11.8	174346	2 AC131356	AC131356 Rattus no
10	299	9.2	144577	2 AC123367	AC123367 Rattus no
11	274.5	8.5	207214	2 AC107643	AC107643 Mus muscu
12	254	7.8	136753	2 CNS08CB6	ALB45344 Oryza sat
13	246.5	7.6	3217	5 AB026623	AB026623 Gallus ga
14	246	7.6	3385	9 AB020652	AB020652 Homo sapi
15	246	7.6	3774	9 AF203032	AF203032 Homo sapi
16	246	7.5	5983	10 AF076623	AF076623 Mus muscu
C 17	243.5	7.5	16832	1 AF070943	AB070943 Streptomy
18	240	7.4	7033	9 AF039571	AF039571 Homo sapi
19	238	7.4	2313	9 BC014185	BC014185 Homo sapi
20	238	7.4	9551	6 AR076233	AR076233 Sequence
21	238	7.4	9551	6 I39845	I39845 Sequence 93
22	238	7.4	9551	9 HMTTRHXL	L09190 Human trich
23	237	7.3	20293	1 SCD20	AL392148 Streptomy
24	236.5	7.3	1222	5 AF239885	AF239885 Gallus ga
25	236	7.3	7033	6 AX010193	AX010193 Sequence
26	234	7.2	2167	9 AB071115	AB071115 Macaca fa
27	234	7.2	8469	9 HSA7041	AJ007041 Homo sapi
28	234	7.2	177607	2 AC119484	AC119484 Rattus no
29	234	7.2	303976	2 LMFICHR15	AL160371 Leishmani
30	233.5	7.2	2800	9 AB040976	AB040976 Homo sapi
31	232.5	7.2	5858	3 AF350276	AF350276 Nepila m
32	231	7.1	150366	2 AC119064	AC119064 Sus scrof
C 33	231	7.1	151578	9 AL589986	AL589986 Human DNA
34	230.5	7.1	29641	3 AC005802	AC005802 Leishmani
35	230.5	7.1	110000	3 AC125735_0	AC125735 Leishmani
36	230	7.1	22970	9 H51247F6	Z68279 Human DNA s
37	229.5	7.1	3621	9 EC005847	EC005847 Homo sapi
38	229.5	7.1	139517	9 AC105446	AC105446 Homo sapi
C 39	228	7.0	10091	1 AF135398	AF135398 Therius b
C 40	228	7.0	146506	8 OSJN00206	AF135397 Oryza sat
41	227.5	7.0	8459	9 AB006626	AB006626 Homo sapi
C 42	227.5	7.0	22693	9 H5PLEB1S3	U63610 Human plect
43	227	7.0	177877	9 AL450472	AL450472 Human DNA
44	226.5	7.0	15534	9 AB051895	AB051895 Homo sapi
45	226	7.0	9121	4 OCCTRICHA	Z19092 O. cuniculus

RESULT 1

ALIGNMENTS

AL590963/c
LOCUS AL590963 196724 bp DNA linear ROD 05-Apr-2002
DEFINITION Mouse DNA sequence from clone RP23-395E10 on chromosome 11,
complete sequence.
ACCESSION AL590963
VERSION AL590963.11 GI:20068458
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Lovell, J.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:16304739.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-395E10 is
from the RPCR-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
FEATURES
source location/Qualifiers
1..196724
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-395E10"
/clone_1lb="RPCI-23"
BASE COUNT 49842 a 49408 c 48640 g 48834 t
ORIGIN
Alignment Scores:
Pred. No.: 3.53e-43 Length: 196724
Score: 1757.50 Matches: 440
Percent Similarity: 29.53% Conservative: 8
Best Local Similarity: 29.00% Mismatches: 10
Query Match: 54.28% Indels: 1059
Gaps: 4
US-10-054-935-2 (1-614) x AL590963 (1-196724)
QY 1 MetThMetArGSeRaLaVaLPhelYsAlaAlaAlaProAlaGlyLysAnProGlu 20
Db 62746 ATATACCATGAGATCCGACGTCTCAAGCGCGCGCCCTCCGCGGCAACCCCGAG 62687
QY 21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyLysProGluAspGluProGlyAla 40
Db 62686 CACCGACTGACTACGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62627
QY 41 AlaGluAlaHisPheLeuProArGHisArgLysLysGluProGlyProProLeuAla 60
Db 62626 GCGGAGGCCATTTCCTCCCGCGGCACTCAAGAGAGCGGCGCGCGCGCGCGCG 62567
QY 61 SerSerGlnGlySerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79

|||||
Db 62507 TCTTCCAGGGCGGAGCCCTCGCTCCAGCGGCTGCGCGCGCGCGCGCGCGCGCG 62507
QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnLysSerPgiYser 99
Db 62506 GGCTTGTACTCCAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62447
QY 100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyLysGlyLysProAla 119
Db 62446 GTGCCCTTGCCCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62387
QY 120 AlaAlaGlyAlaGlyLysSerProArGProLysTyrGlnAlaValLeuProLysGln 139
Db 62386 GCACCGCGCGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62327
QY 140 GlySerLeu---ValAlaAlaAlaLysGlnProThrProThrAlaGlyLysGly 158
Db 62326 GGCTCTATCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62267
QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro 178
Db 62266 GCGGCTGCCCTGACCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62207
QY 179 GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyArg 198
Db 62206 GGGCCACACACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62147
QY 199 TrpLysSerMetArgLysSerProLeuGlyGlyGlyLysSerGlyLysSerGln 218
Db 62146 TGGAGAGATTAAG 62087
QY 219 AlaAlaCysLeuLysGlnLeuLeuLeuGlnLeuAspLeuLysGlnGlnGln 238
Db 62086 GCGCGCTGCTCAACAGATCTTGTCTGCTCAATGAGACTCATCAACAGCAGCAG 62027
QY 239 GlnLeuGlnAlaLysGlnLysGlnLysGlnLysGlnLysSerLysArgAspThr 256
Db 62026 CAGTTCGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61967
QY 256 ----- 256
Db 61966 GAGGGGTAATCTGCATTAGAGGATGGAGAGAGAGAGAGAGAGAGAGAGAG 61907
QY 256 ----- 256
Db 61906 CTGTAGAGAGCAGAGGTCCCGCTCGGTAGAGGTTAGAGAGAGTGGCAGCTAAAG 61847
QY 256 ----- 256
Db 61846 AGTGCTTAGATGTGGGTCTTAGAGAGCTGACATCACTGACAGCTGGGTGGAAGA 61787
QY 256 ----- 256
Db 61786 AGGGTTAAAGGCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 61727
QY 256 ----- 256
Db 61726 AAAGGTTAATTAATAACAGACAGCCACACTGTGTGGAAGAGAGCTTTAGAGTGAAG 61667
QY 256 ----- 256
Db 61666 ACCTGATTACGAGGGGTTAAAAATGATAGAGGTGGGGTGCAGATGTGTGTGTGTGT 61607
QY 256 ----- 256
Db 61606 GGTGTATTACAGCGCTGTGCTTAGCTGTGATGTGATGTGAGAGAGAGCTTAACAGCTGTT 61547
QY 256 ----- 256
Db 61546 GCGGTAACACCGCTGTGTGATCTTTAGACAGAGAGGTTGTCTATAGTCTTTTGTGCGGT 61487
QY 256 ----- 256

Db	61486	AGAGGGTAAGATGCTGTACGTGGAACTCTGTTTAACTCTCCCTAAATGAGATCACTGC	61427
OY	256	-----	256
Db	61426	CTTTGATTTACGAGAGTTCTTAGTAAAAATCGAGTTATAAACACTACCGCTGACCA	61367
OY	256	-----	256
Db	61366	CGTCCCTTGACTAAGCTGGCAGATTTTGAGACATAAGCTTATTCGATGATTCTGTAAC	61307
OY	256	-----	256
Db	61306	CCTTCCACCTCCAGACCTGTTTTGAACTTGAGGTAATTAGAAAAATTGTTGTTAATT	61247
OY	256	-----	256
Db	61246	CTTTGTACTCTGGAAGAATGACATGACCTCATAGGCTTGCTGCATGAGCAGACACTGGGC	61187
OY	256	-----	256
Db	61186	ACACATGTTAAGCACACTCCCTCACACTAGGGCCGTGGGGTGGTGGTCTACAGTG	61127
OY	256	-----	256
Db	61126	GTAACATACCTTTTAGACGCTAAAAATAAACATGGCTTCTACAGTGGCTCAGGCCCTG	61067
OY	256	-----	256
Db	61066	CTCCCCCTCCCTCCTCCTGGCTCCTTACTCTGGGGCTTCAGCAGCACAGGCTGAG	61007
OY	256	-----	256
Db	61006	CATGGAGCAGAAAGCTGTTCATTTTGTGCTCTTAATAAGTCGCCATCAGGCCCTC	60947
OY	256	-----	256
Db	60946	TGTTTGATTGGAGATGCTCTTCGAAGGTGAACAGAGCTGGGGCAGGCTTGTGAA	60887
OY	256	-----	256
Db	60886	ATACCTCAGTGTTCCTTTTGGAACTACTCAGATGAGTGGCACTGCTCAGCACTGTA	60827
OY	256	-----	256
Db	60826	CTGCAGCTACGGTGTGATGAGCCCTGCCGTAAAAACAACAATATCCTTCCTTATGGG	60767
OY	256	-----	256
Db	60766	AAGGGGTGGAGAAGAGATCAGTAAAGAGAAAGAAATGAAGTATGTACCCACAGA	60707
OY	256	-----	256
Db	60706	AACATGCTAAGTAATGCTAAGGCTGTGAATCAGAACACAGGTGACAGAGTGGCTGCCA	60647
OY	256	-----	256
Db	60646	GGGGGCGGGGAGACACTACTGCAGCTGGAGGGTGTGTGAGTTCTTGAGGATGACCTT	60587
OY	256	-----	256
Db	60586	CTTTGGCTCCTTCATAGGCTTTTCAAGTAGGAAAAAGTGGGTGAGTTAATGCTAAG	60527
OY	256	-----	256
Db	60526	CATTGGTCTAGCTTGGCATTCACGGCTTGTGATGTACTACAGTCTTAAGCCCTGCTTA	60467
OY	256	-----	256
Db	60466	GGCCTGCTTATTCTAGTCATTTGTTGGCCTCTGCTGAAGTCATGCCATGGCAGTAAG	60407
OY	256	-----	256
Db	60406	AGAGCAGACTGACGTCACTCACTAACAAGCAAAAGCTTGTCTTGTGCTATAAGCAG	60347

QY	256	-----	256
Db	60346	CACAGTCACCCTTTTGTCAAACTGAGTTAATACCAAGTTTCTGATGACAATFAGACTTCC	60287
QY	256	-----	256
Db	60286	ATTTTGTTTGATTTTTGGTTCCTCTTAATTGAATGTAAITTTGATAAATATTAGAT	60227
QY	256	-----	256
Db	60226	AATATTTTCCCGTTATCTCCAACACAGAAGCATCTAATATGTCATGTGTGGAT	60167
QY	256	-----	256
Db	60166	TATATTACTGGCGTCGATTAATATCTTTTCACTGATATGTGTTTAGAATTACTCCTTC	60107
QY	256	-----	256
Db	60106	AAATACGAATCATACATTTTGTATGTGTAAGTATCCCTPAGGACGCTGTGAGCC	60047
QY	256	-----	256
Db	60046	CTTGATTCCTCTGCATTTGTGCTTTAAACATTCACGTGTCTAGAGTTTGGAGTAA	59987
QY	256	-----	256
Db	59986	CATGAGCAGATAGATTTGGACACCGTAGACACTTAAAAGAACACAGGCTGAGAA	59927
QY	256	-----	256
Db	59926	TGTAGCTCAGTGTGCTGATGCTTAGAGTGTGTCAGCACGCGTCAGAGGCCCTGATT	59867
QY	256	-----	256
Db	59866	CCATCCCTAGCAGCACAGACGACGACCAACAGAAACAGATTAAATTGAGGGGCTCC	59807
QY	256	-----	256
Db	59806	CAGAAAGACTTGAGCACAAGTGCGAGAGCGCTTGCCATTAAGCACTCCGCTCTGCC	59747
QY	256	-----	256
Db	59746	AGTGGAAAGGCGCTCCCTTCAGAAAGAGCTTCAGCGTCCCACTGGGAGGGCTCC	59687
QY	256	-----	256
Db	59686	TTCAAGAAAGAGACTTCAGGCTCCCTTACTCTTCTGCCATCCAGACCAACTCTACAA	59627
QY	256	-----	256
Db	59626	TACATTAAGTTGTCTTTGCCCTTCATATGGGTTTAAAGGAAGCTATTCAGATTCGTGCTG	59567
QY	256	-----	256
Db	59566	CCAGTTCAAGATAAAAGTTGAATTGTAGGACCATAGCGGCACGAGAAGATCGCTGAT	59507
QY	256	-----	256
Db	59506	TCCATAATGGGCTAAGACCGCGGCTTAACCTTGTGCTGAAGTCTGTACTCTGTACCT	59447
QY	257	-----LeuLeuAlaArgIleGluArgMetGlnuArgHmetGlnLeuValLysLys	273
Db	59446	TCCCTCCAGCTCCCTTGGCTCGGATGAAACGTATGAAAGGGGAGTGCAGCTGGTGAAGAGG	59387
QY	274	AspAsnGluLysGluArgHisLysLeuPheGlnIleYTrGlnurHrgLuuArgGlnuGln	293
Db	59386	GATTAACGGAAGAAAGCAACAGCTGCTCCAGGGCTATGAOCTGAAGAGAGAGAGAA	59327
QY	294	ThrGluLeuSerGluLysIleLysLeuGlnCysGlnProGluLeuSerGluTrpSerGln	313
Db	59326	GCAGAGTTGTCGTGGAAAAATTAATTTGGACCGCCAGCGGAGCTTTGCGAGACTTCCAG	59267

QY	334	ThrlaupProPolysProPhSeSerCysglYarGserGlyLysglYH:sls	-----	330
Db	59266	GCTCGCCTTCCAAAGCCTTCTCATGTGGCCGAGTGGAAAGSACACAAAAGGTGCT		59207
QY	330	-----		330
Db	59206	GGTCACCTTATCTGTACTGMAAAAGATGAGGAGCGCTAATGTGATGAAATATAG		59147
QY	330	-----		330
Db	59146	CTTGGGGTTCGTGCATTGGCATTATACCAACCAATACCTGAATGAGAAGCTGTATTTC		59087
QY	330	-----		330
Db	59086	TAGGGGATGTGGCTTACCCCTGTATACACAGCAGCTTAGAGTGTAAAGCAAGATCT		59027
QY	330	-----		330
Db	59026	GTTGTGAGTTCAGAGCTGAGCCTGGCAGTAGGGAACCTGTTTTTACAAACAAAAGCTTC		58967
QY	330	-----		330
Db	58966	CAGGTTCTGTGTGAGAAATGCTTGAAACCAATTTCTTGTCACAGTACTAAGTACTC		58907
QY	330	-----		330
Db	58906	GGGTGCCAGCTACATCCCAAGAGCCTGCTGTATACTTTGAAGCCTGACCTGTGTGC		58847
QY	330	-----		330
Db	58846	CCATGGCTTCCAAACGACAGTAGATTTTAGTACCTGGATACCTCTCTTGCCCTCAAA		58787
QY	330	-----		330
Db	58786	TATATCAAAATCAATTTTATTGAAATCAATTATTTTTGTTTTGTTTGTGTTTTTG		58727
QY	330	-----		330
Db	58726	AGACGGGTTCTCTGTAGACACAGCGTGGCTCGAATCAGAAATCGCTGCTGTGCC		58667
QY	330	-----		330
Db	58666	TCCCTAGTGGGATTAAGGCGTGGCCACACAGTCCCGGCCCTGGAATCAATTATTT		58607
QY	331	-----	ArgLysSerProheGlySerThrGluArgLys	341
Db	58606	TTTATTAACATTTTTTTCTACTAATTAAGAAAAACCCATTTGGAAATACAGAAAGAAAG		58547
QY	342	ThrProValLysLysLeuAlaProGluPhSeSerLysValLysThrLysThrProLysH:sls	-----	361
Db	58546	ACTCGCTTTAAAAAGCTGGCTCTGAAATTTTCAAAAGTCAAAACAAAACCTCTTAAGCAC		58487
QY	362	SerProLLeLysglLugLInuProCysglYSerLeuSerGluLuthrValCysLysArgGluLeu	-----	381
Db	58486	TCTCCCAATTAAGAGGAAACCTGTGGTTCATATCAAGAAACGTGTGTAAACGTGAATTG		58427
QY	382	AlrSerGlnGluThrProGluLysProArgSerValAspThrProProArgLeuSer	-----	401
Db	58426	AGGAGCCAAAGAAACCCAGAAAAAGCCCCGCTCTTCAGTGAATACCCACCAAGACCTCG		58367
QY	402	ThrProGlnLysglYProSerThrHisProLysGluLysAlaPhSeSerSerGluLeGlu	-----	421
Db	58366	ACTGCCCAAAAGGAGCCACACCCACCCACCAAGAGAAAGAACCCCTTCCAACTGAGATGGAA		58307
QY	422	AspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTPrHisGlnProProPo	-----	441
Db	58306	GATTTGGCCGTACCTTTCACACAGAAATGTATTGTGTGCTGGACACAGCCTCCCCCA		58247
QY	442	SerProLeuProLeuArgLysSerSerProLysLysGluLuthrValAla	-----	458
Db	58246	TCAACGTTACCATTTACGGGAATCCTCTCCCAAAAGAGAGAGACTGTATACA		58196

	AC068669	180017 bp	DNA	linear	HTG-16-FEB-2002
LOCUS	AC068669	Homo sapiens chromosome 17 clone RP11-749I16 map 17,	*** SEQUENCING		
DEFINITION	IN PROGRESS ***; 16 unordered pieces.				
ACCESSION	AC068669				
VERSION	AC068669.4	GI:18677458			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 180017) Biren,B., Linton,L., Nusbaum,C. and Lander,E. Homo sapiens chromosome 17, clone Rpl1-749I16 Unpublished 2 (bases 1 to 180017)				
AUTHORS	Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barua,N., Bastien,V., Bedal,F., Boguslavsky,I., Boukhgalter,B., Brown,A., Burrell,G., Campopiano,A., Castle,A., Chepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Donino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Laocque,K., Lamazeres,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,C., Mcpheeters,R., Melidim,T., Menaus,L., Mihova,T., Miranda,C., Mianga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,C., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer.A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-May-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (bases 1 to 180017)				
REFERENCE	Biren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barua,N., Bastien,V., Boguslavsky,I., Boukhgalter,B., Brown,A., Camarta,J., Campopiano,A., Chang,J., Chararo,B., Chepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., Laocque,K., Lamazeres,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,A., Macdonald,P., Major,J.T., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Melidim,J., Mihova,T., Mianga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norby,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riback,M., Riley.R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauhan,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Feb 15, 2002 this sequence version replaced gi:11597011.				
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR				

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L10045
Center clone name: 749_L16

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 7782: contig of 7782 bp in length
7783 7882: gap of 100 bp
7883 15614: contig of 7732 bp in length
15615 15714: gap of 100 bp
15715 17880: contig of 2166 bp in length
17881 17980: gap of 100 bp
17981 26841: contig of 8861 bp in length
26842 26941: gap of 100 bp
26942 36150: contig of 9209 bp in length
36151 36250: gap of 100 bp
36251 62842: contig of 26592 bp in length
62843 62942: gap of 100 bp
62943 72161: contig of 9219 bp in length
72162 72261: gap of 100 bp
72262 82180: contig of 9919 bp in length
82181 82280: gap of 100 bp
82281 87039: contig of 4759 bp in length
87040 87139: gap of 100 bp
87140 104295: contig of 17156 bp in length
104296 104395: gap of 100 bp
104396 129621: contig of 25226 bp in length
129622 129721: gap of 100 bp
129722 144612: contig of 14891 bp in length
144613 144712: gap of 100 bp
144713 149573: contig of 4861 bp in length
149574 149673: gap of 100 bp
149674 153213: contig of 3540 bp in length
153214 153313: gap of 100 bp
153314 164194: contig of 10881 bp in length
164195 164294: gap of 100 bp
164295 180017: contig of 15723 bp in length.

FEATURES

Location/Qualifiers

1..180017
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"

BASE COUNT 43238 a 43942 c 45220 g 46021 t 1596 others
ORIGIN
/clone="RP11-749116"
/clone_id="RP11 Human Male BAC"

Alignment Scores:

Pred. No.: 2,31e-34 Length: 180017
Score: 1460.00 Matches: 463
Percent Similarity: 20.07% Conservative: 1
Best Local Similarity: 20.03% Mismatches: 7
Query Match: 45.09% Indels: 1842
DB: 2 Gaps: 4

US-10-054-935-2 (1-614) x AC068669 (1-180017)

QY 1 MetThmetarSerAlaValPheLysAlaAlaAlaAlaProAlaGlyAsnProGlu 20
Db 66354 ATGACCATGAGATCCGGGTGTTCACAGCGCGCGCCCTGCGGCAATCTGAG 66413
QY 21 GlnArgLeuAspTyrGluArgAlaAlaAlaLeuGlyCysProGluAspGluProGlyAla 40
Db 66414 CAGCGACTGACTACGAGCGGCTGCGGCTGCGGCGCGCGCGAGAGAGCTGCGGCG 66473

QY 41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProLeuAla 60
Db 66474 GCCACACCCACTTCTCCCCCGGACCGTAAGCTCAAGAGCCGGGCGCCCTGACC 66533
QY 61 SerSerGlnLysLysSerProAlaProSerProAlaGlyCysGlyLysGlyArgGly 80
Db 66534 TCTCCACAGGCGGAGACCCCGGCTTCCCGGCGGTGCGGCGGCAAGGCGCGGCG 66593
QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGln 100
Db 66594 TTGTTACTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66653
QY 101 ProLeuProCysProProProAlaThrLysGlnAlaGlyLysGlyGlyProAlaAla 120
Db 66654 CCTTGGCCCTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66713
QY 121 AlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGly 140
Db 66714 GCGGAGCGCGGCTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66773
QY 141 SerLeuValAlaAlaAlaLysGluProThrProThrAlaGlyAspLysGlyAlaAla 160
Db 66774 TTCTCTCGGCG 66833
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProProLeuProGlyPro 180
Db 66834 TCCCCCGCTGCCACCGCGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66893
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTyrLys 200
Db 66894 CCACCCCTCG 66953
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 220
Db 66954 AGTATGAG 67013
QY 221 CysLeuLysGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 240
Db 67014 TGCTCTCAACAGATCTTCTGCTCTCAATGGACTCATCGAACGACGACGACGACTG 67073
QY 241 GlnAlaLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 260
Db 67074 CAGGCCAAGGAAAG 67133
QY 256 ----- 256
Db 67134 TTAAATCTGCATTCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67193
QY 256 ----- 256
Db 67194 GAGGTAAAGCACCCTCGGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67253
QY 256 ----- 256
Db 67254 TAGCGCTGGGTCTCTGGGAGCGAGCGTCAACAGACAGCTTGAGTGAAGGAGGTAA 67313
QY 256 ----- 256
Db 67314 AGGCAACCTCTCAGGGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67373
QY 256 ----- 256
Db 67374 ACAGCCCAAGAGCTGTGGGAAAGAGCTTTAGTGGAAAAAGAGAGAGAGAGGTAA 67433
QY 256 ----- 256
Db 67434 AAATGAAGAGGTGGGTGTAGGTGTGTCTTATGAAGGTAAATTAAGAGCTA 67493
QY 256 ----- 256
Db 67494 TTGCGTTAAGAAATGACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67553

QY	256	-----	256
Db	67554	TGCTGCAGCATCTTACATAGAGAGGCGATGAGCCCTTTAAGTGGCACAAGGGGTAGAGGT	67613
QY	256	-----	256
Db	67614	GATTTACGTGGAATCCTATTTAGTTTTCTTAATGATATTAGAGAAATCAGCTTTG	67673
QY	256	-----	256
Db	67674	ATTTCAGACAGCTTTTAATAGAACCTGAGTTCTAATACATGACCCAGTACAGGTCCC	67733
QY	256	-----	256
Db	67734	TTTTATAGCATGSGTTTGAAACTGSAAGATTTTAGGGCATCAGGGGAAGTAAGCTTA	67793
QY	256	-----	256
Db	67794	TTAAATGATTAATATAGACCTCCACATCCTTATTTAGAACTCTTAGGAGTAGAAGATC	67853
QY	256	-----	256
Db	67854	GTTTATATTCTTCTTACTCTGAGAGCCTTGACAGACAGGCTTGTTTTACTGCAGCA	67913
QY	256	-----	256
Db	67914	GTAGTTAAGGGCATCTCGGTTAAACAATATCTTCAGATATGAGGGAAGCTGGGCTG	67973
QY	256	-----	256
Db	67974	GGGGGTATTACTGTGCTAGCATCTTACAGCTGCCTTAAATAATACATGTTTCTACAA	68033
QY	256	-----	256
Db	68034	TTGTTGATTCAGGCTGCAGACAGTCTCTCTCCCTCTGCTCCCTCTTCTCCTCCCTCC	68093
QY	256	-----	256
Db	68094	CTGCTCTTCCCTTCTCTCTCTCTCTCTCTCTCCCTCCCTCCCGCGTCTCATCTCTC	68153
QY	256	-----	256
Db	68154	CCCTCTCTCCTTCTCTCTGCGGTGCTGCGCCTAAGACTGAGCAAGGAGCAAGCG	68213
QY	256	-----	256
Db	68214	CTATTCATTTTGTGTCTTCCAAATGAATCTCATCGGGTGGCTGTTTATTGGAAATG	68273
QY	256	-----	256
Db	68274	CATTCTCAAGTGATGGAGCTGGGCGAGTCTCATTTAAATATCTCATGTGTGCA	68333
QY	256	-----	256
Db	68334	TTTGTGGAAGTTACTCAACATGAGGATGGCAGTCTGACAGTGTACTGCAGTCAAGGGT	68393
QY	256	-----	256
Db	68394	GTCATGCTACCGTATTTACACACATCATATCTCTTATGAGGGGAGTGGGAAGAAGG	68453
QY	256	-----	256
Db	68454	ACCAGTGAAGAGAAAAAATCAAGGAAATGTCCACCACAATAATGTGTTAAGGCTGTG	68513
QY	256	-----	256
Db	68514	AATCAAAACATATACAGAGAGATGAATGAGAGTATGCTGATAGGGGGTTCCTGCTG	68573
QY	256	-----	256
Db	68574	ATTGTGAGGTTCTGCTCAGTTCTGTAGGTAATGGCCCTTTGCAGGCTTCTTAAGTCA	68633
QY	256	-----	256

Db	68634	GGAAAGGGGGCTAGTTAAGATAGAGTCTGACGTGAACTTGATGACCTGGTACCTGCTTCG	68693
QY	256	-----	256
Db	68694	TTGTGATTTGCAATCTTAATATTGCTTAAGCTTGTTCACTTAATTCAGTTGTTGTGG	68753
QY	256	-----	256
Db	68754	CCTTGTCTGAAATCTATGCAATGATGATGAAGCAGCGTTAAACATCAGTCTAAC	68813
QY	256	-----	256
Db	68814	CCAGATAAATTTCTGTAAATAGAGTAAAGTAGACACACACTTTCTCAAAACTAAACT	68873
QY	256	-----	256
Db	68874	AACCAAAAGTTTCAATACTAATAGAGTCCATTTCTGTAGTTTTTACTCTTCCT	68933
QY	256	-----	256
Db	68934	TTTAAGTAAAGCTAAGGAAGCTTGATTTTAATAAATATTAGATGATAGTAAATTT	68993
QY	256	-----	256
Db	68994	TTTGTGTACCTCAATTAAGGAATGTGTTATGTCATGATTTGGAAATTAGGTA	69053
QY	256	-----	256
Db	69054	CTGATTTCTATTTTCCAAATGAATGTTCATTTGTTTAAAGTACCCCTCCACACATA	69113
QY	256	-----	256
Db	69114	CCCCCTTCAATAAAGACTGCGCATTCGTCGTCAGTGTAGAGCTTAATTCCTTCA	69173
QY	256	-----	256
Db	69174	ATGTACTGTTTGACCTTGCTGTTGTGTATAGAGCTTTGGAGTAGAAGTAGAGGATAG	69233
QY	256	-----	256
Db	69234	AATTTGGACAATCTCAGAGGCGCTGTAAAGATCATGTAGGCTGGCGGTGCATCAGC	69293
QY	256	-----	256
Db	69294	CTGTAAATCCAGCAGCTTTGGAGGCGCTACGAGGTGACACGAGTCAAGATAGAGA	69353
QY	256	-----	256
Db	69354	CCATCTTGGCAACATGTGAATCCCTGTCTACTAAAAATCAAAAAATTACCGCGTGT	69413
QY	256	-----	256
Db	69414	GGTGCGGCGGCGCTGAATATCCAGCTACTCGGAGGCTGAGCAGGAGGCGAGGTTGC	69473
QY	256	-----	256
Db	69474	AGTAGGCATTTGCATCCAGCCTGGGCAACAAAGCAAAACTCCCTCAAAACAAAAAC	69533
QY	256	-----	256
Db	69534	AAAAACAACAAACAAACATGTATACATTAATTTCTGGAACCTCCGGAACCAACTTGGG	69593
QY	256	-----	256
Db	69594	GCCACACAGGTAGCAGTAGCAGAGACCAGTCATTAAAGGCACTTTACGGCCCTTCCA	69653
QY	256	-----	256
Db	69654	GTGGAGAGGATCTCCCTACAGAGAAATGAACCTCAAAATTACCCCATGCTCTTCCGATCT	69713
QY	256	-----	256

Db 69714 AGGTTAGCCCTTAATAATTTAACAGTATACCTTCTCTCTATTTGTTTAAGGAAACT 69773
QY 256 ----- 256
Db 69774 GTTCAGATTAAATGCTCTCTATTTAAAGTTAGTTGATGTGTCATAGAAAGAGCA 69833
QY 256 ----- 256
Db 69834 GAAGACTATACTGATTCCTAAATGGGCTAAGTATCTGTGCTTATTTTATGTGTT 69893
QY 257 ----- LeuLeuLlaargllegluatgmetluarg 266
Db 69894 AAGTCTCAATTTGCTACTCTCTCTTTTAACTCTCTCTGATGATGAACGATGGAAG 69953
QY 267 ArgmetGlnLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTyr 286
Db 69954 GGGATGAGCTGGTAAGAAAGATACGAGAAAGAGCAAGCTCTTCACAGGCTAT 70013
QY 287 GluThrGluLuarGluGluLuarGluLeuSerGluLysIleLysLeuGluCyGlnPro 306
Db 70014 GAAACTGAGAGAGAGAGAGAAAGAGAGCTATCTGAGAAATTAACGAGAGTGCACCG 70073
QY 307 GluLeuSerGluThrSerGlnThrLeuProProlLysProPheSerCysGlyArgSerGly 326
Db 70074 GAGCTTCCGAGACATCCAGACTCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGA 70133
QY 327 LysGlyHisLys ----- 330
Db 70134 AAGGGACATAA -AAGGTGCTGTAACTTTGATGAGAGCCCTGTTACAGATGAT 70192
QY 330 ----- 330
Db 70193 TGAGAGTAGAGATTGGGATTATGATTTGTTATAGCAATCATAACTTGTTAGAA 70252
QY 330 ----- 330
Db 70253 GTCTTATGCGAGGCGGGGTGTGTGCTCACACTGTAATCCAGCACTTGGAG 70312
QY 330 ----- 330
Db 70313 GCCAAGCGGGGTGATCATGAGTCAAGATAGAGACCATCTGGCCAACTTTTGAAA 70372
QY 330 ----- 330
Db 70373 CCCCCTCTACTATAAATATAAATAATTAGCCGGCGGTGATGGCGGCGCTTACTCC 70432
QY 330 ----- 330
Db 70433 CAGCTACTCGGAGGCTGAGGAGAGATCATCTTGAACCGGAGGCGAGGTTGAGT 70492
QY 330 ----- 330
Db 70493 GAGCGGGGATTGCTCACTGCACTCCAGCTTGGCAAGAGAGGAGACTCCGTCAAAAAA 70552
QY 330 ----- 330
Db 70553 AAAAAAAGAAAGTCTTCTATGCTGAGAACTATCTTGAACATTTCTTATGACACGTGA 70612
QY 330 ----- 330
Db 70613 AAAGCTGTTAGAGATTTTGTGTACAGTAATATATGAAATGAGCTTACATCAT 70672
QY 330 ----- 330
Db 70673 AAGGAATGGCAATAAGTATGACAGCTGAGTTCTTAAACTATAGCAGTTCACTTCAA 70732
QY 330 ----- 330
Db 70733 GTTAGGAAGTGTAAATTCCTTCAATGTGTGCTGATTTGTAGACATAGAAAAGTAAT 70792
QY 330 ----- 330
Db 70793 ACAGGCAAGCGGTGTGCTCATGCTGTAAATCCCTTTGAGAGGCAAGGTGGGAAAGA 70852

QY 330 ----- 330
Db 70853 GTATGAAAGTATGAAAGTGGGAAGAGAGGATCGCTTGCCCGAGAGTTTGAGACCA 70912
QY 330 ----- 330
Db 70913 GCCTGGCAATFATGAGACCTCATTTTACAAAAAATCCAAAAATTAGCGGGGCTGG 70972
QY 330 ----- 330
Db 70973 AGTATGCACTGTAGTCCAGCTACTCAGAGGTTGAGGCTACAGTGAAGCTTTACTGC 71032
QY 330 ----- 330
Db 71033 ACCACTGACTCCAGCTGGGTGACGAGTATACCTGTCTCAATTGAAAAAATAAAA 71092
QY 330 ----- 330
Db 71093 AAAAAGTAAATGAAGTGTCTCTGACTCAAAAAAATTAACCCAACTGTAAAGTAG 71152
QY 330 ----- 330
Db 71153 GCATACATTTTGAAGACTCAAAATCAAAATGCAAGTAAATAGTAAACACACCTG 71212
QY 330 ----- 330
Db 71213 ATATAACTAAGCTTTTGTGTTGTGATAGAGGTCTCTCTGTACACCAAGGCTGG 71272
QY 330 ----- 330
Db 71273 GTGCAAGTGTGCCAATATGCTCCTCACTTCACTTCCACCTGTGAGCTCAAGTATCTC 71332
QY 330 ----- 330
Db 71333 CTACTCAGCTCCCATGTAGCTAGACCATAGGCACTCACACATGTTGGCTAATTTT 71392
QY 330 ----- 330
Db 71393 TTAGTTTTGTAGACAGGCTCTCACTTTGTTTCCAGGCTGATCTCAAACTCTGGGC 71452
QY 330 ----- 330
Db 71453 TCAGTATCTCTCCGAAGTGTGGATTTACAGACATTAACATAAGTTTGTAAATCA 71512
QY 330 ----- 330
Db 71513 GGGATATTAAGTTATTTCTGTGTTAGAAATTTGAACTTTAAGTAAATCAGTCTTGA 71572
QY 330 ----- 330
Db 71573 AAAAGAAATATAGCCGGGTGCGGTGCTCACGCTGTAAATCCAGCACTTTGGGAGCG 71632
QY 330 ----- 330
Db 71633 AGGGGGTGAATCAGAGGTGAGAGATTCAGACCATCTGCTTAACACAGGTGAACCT 71692
QY 330 ----- 330
Db 71693 GTCTTACTTAAAAATAGAAAAATTAGCCGGGTGTGTGCAATACACTGTAGTCCAGCTA 71752
QY 330 ----- 330
Db 71753 CTTGGAGGCTGAGGAGAGAAATGCTTGAACCCAGAGGTGAGGTTGCAATAGCAG 71812
QY 330 ----- 330
Db 71813 AGATCGACCACTGCACTCCACTCTGAGGTGACAGAGTAGACCTCAAAAAAGTATTAATA 71872
QY 330 ----- 330
Db 71873 TTTAGTTGTGAAGAGAAAGAGATGACTTTCTTGGATGTGTGGGGAAAGAGGA 71932

OY	330	-----	330	-----	330
Db	71933	AGAAAGAGAGGCTAGAGATGCCAGCAGCACTAAGAAAAAGAAATTAAGGTTACAAT	71992		
OY	330	-----	330	-----	330
Db	71993	TAATGTGTACAGGCTTGGGGAAAAAGAAAGACTGATCTCTGATTACAGAAGTAG	72052		
OY	330	-----	330	-----	330
Db	72053	GTGCACCTAATGTGGGTTGTAGTGTACATAAATGTGGCTAAAGACCTTGAATATCTT	72112		
OY	330	-----	330	-----	330
Db	72113	TTTTTTTTTTTTTTTTTTTGTGAGACAGAGCTCACTGTCGCCANNNNNNNNNNN	72172		
OY	330	-----	330	-----	330
Db	72173	NN	72232		
OY	330	-----	330	-----	330
Db	72233	NN	72292		
OY	330	-----	330	-----	330
Db	72293	TATTCTTAAGCATGGAATTCCTGTTCACAATTTACTGGCCAGGTGCAGTGCATGTC	72352		
OY	330	-----	330	-----	330
Db	72353	CGGTAATCCAGCACTTGGGAGGCGTAGGTGGTGGATCACCTGAGTCAGGTATTTGA	72412		
OY	330	-----	330	-----	330
Db	72413	GACCACTGTGACCAATATGGTGAACCTGCTTTATTAATAATACAAAAAAATTAACC	72472		
OY	330	-----	330	-----	330
Db	72473	AGCAGTGGTGTATGCTGTAACTCCAGACTCTAGAGAGCTGAGATAGAGAAATTGC	72532		
OY	330	-----	330	-----	330
Db	72533	TTGAACCGGAGGACAGATTTGCAGTGCACGACATGCCACTGCACCTCAGCTTGG	72592		
OY	330	-----	330	-----	330
Db	72593	GTGACAGACAGACTCTGTCTCAAAAACAAAAACAAAAAATTTACTATGG	72652		
OY	330	-----	330	-----	330
Db	72653	GCCATTATTTATATCTGCTAAGCCTGGGCTTCTGCTCATGACTGTTAAATGAGCACC	72712		
OY	330	-----	330	-----	330
Db	72713	AGACATTTAGGTACCTGAATACTTGTGCATTTTGGGGCAAAAGAACAGTTATCTTTTG	72772		
OY	330	-----	330	-----	330
Db	72773	CTCCAGATGTGTTCAAAATTAGATTTTATTCAGTGTCAATAAATTTTAAATGTTTCT	72832		
OY	331	-----	331	-----	331
Db	72833	TCCCTATCTAATAGGAATCCCATTTGGAAAGTGAAGAAAGAAAGCTCTCTTTAAAG	72892		
OY	347	LeuAlaProGluPheSerLysValLysThrLysThrProLysHisSerProLysGlu	366		
Db	72893	CTGGCTCTCTGAATTTTAAAAAGTCCAAAACAAACTCTAAGCACCTCTCTTTAAAGAG	72952		
OY	367	GluProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnIuThr	386		
Db	72953	GAACCCGTGGTTCCTTATCTGAACGCTTTGAACGTAATGTAGAGCCAAAGAAC	73012		
OY	387	ProGluLysProArgSerSerValAspThrProProArgLeuSerThrProGluLysGly	406		

[illegible]

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193189)
AUTHORS Morley K.C.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 193189)
AUTHORS Morley K.C.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20387269.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GURK
Center clone name: CH230-28N16
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 126440 bases at least Q40
Consensus quality: 140831 bases at least Q30
Consensus quality: 148852 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1554: contig of 1554 bp in length
1555 1654: gap of unknown length
1655 2715: contig of 1061 bp in length
2716 2815: gap of unknown length
2816 3899: contig of 1084 bp in length
3900 3999: gap of unknown length
4000 5367: contig of 1368 bp in length
5368 5467: gap of unknown length
5468 6690: contig of 1223 bp in length
6691 6790: gap of unknown length
6791 8012: contig of 1222 bp in length
8013 8112: gap of unknown length
8113 9566: contig of 1454 bp in length
9567 9666: gap of unknown length
9667 10914: contig of 1248 bp in length
10915 11014: gap of unknown length
11015 12395: contig of 1381 bp in length
12396 12495: gap of unknown length
12496 14050: contig of 1555 bp in length
14051 14150: gap of unknown length
14151 15648: contig of 1498 bp in length
15649 15748: gap of unknown length
15749 16994: contig of 1246 bp in length
16995 17094: gap of unknown length
17095 18274: contig of 1180 bp in length
18275 18374: gap of unknown length
18375 20426: contig of 2052 bp in length
20427 20526: gap of unknown length
20527 22027: contig of 1500 bp in length
22028 22126: gap of unknown length
22127 23257: contig of 1131 bp in length
23258 23357: gap of unknown length

23358 24652: contig of 1295 bp in length
24653 24752: gap of unknown length
24753 26187: contig of 1435 bp in length
26188 26287: gap of unknown length
26288 27567: contig of 1280 bp in length
27568 27667: gap of unknown length
27668 28980: contig of 1313 bp in length
28981 29081: gap of unknown length
29082 30136: contig of 1056 bp in length
30137 30235: gap of unknown length
30237 31632: contig of 1396 bp in length
31633 31732: gap of unknown length
31733 33270: contig of 1538 bp in length
33271 33371: gap of unknown length
33372 35465: contig of 2095 bp in length
35466 35565: gap of unknown length
35566 37201: contig of 1636 bp in length
37202 37301: gap of unknown length
37302 38947: contig of 1646 bp in length
38948 39047: gap of unknown length
39048 40710: contig of 1663 bp in length
40711 40810: gap of unknown length
40811 42463: contig of 1653 bp in length
42464 42563: gap of unknown length
42564 43912: contig of 1349 bp in length
43913 44012: gap of unknown length
44013 45591: contig of 1579 bp in length
45592 45691: gap of unknown length
45692 47620: contig of 1929 bp in length
47621 47720: gap of unknown length
47721 49829: contig of 2109 bp in length
49830 49929: gap of unknown length
49930 51465: contig of 1536 bp in length
51466 51565: gap of unknown length
51566 52865: contig of 1300 bp in length
52866 52965: gap of unknown length
52966 54343: contig of 1378 bp in length
54344 54443: gap of unknown length
54444 55895: contig of 1452 bp in length
55896 55995: gap of unknown length
55996 57574: contig of 1579 bp in length
57575 57674: gap of unknown length
57675 59301: contig of 1627 bp in length
59302 59401: gap of unknown length
59402 61393: contig of 1992 bp in length
61394 61493: gap of unknown length
61494 63075: contig of 1582 bp in length
63076 63175: gap of unknown length
63176 64817: contig of 1642 bp in length
64818 64917: gap of unknown length
64919 67403: contig of 2485 bp in length
67404 67502: gap of unknown length
67503 69777: contig of 2275 bp in length
69778 69877: gap of unknown length
69879 71982: contig of 2105 bp in length
71983 72082: gap of unknown length
72083 74289: contig of 2207 bp in length
74290 74389: gap of unknown length
74390 76937: contig of 2548 bp in length
76938 77037: gap of unknown length
77038 79003: contig of 1966 bp in length
79004 79103: gap of unknown length
79104 81519: contig of 2416 bp in length
81520 81619: gap of unknown length
81620 84864: contig of 3245 bp in length
84865 84964: gap of unknown length
84965 87539: contig of 2575 bp in length
87540 87639: gap of unknown length
87640 89437: contig of 1798 bp in length
89438 89537: gap of unknown length
89538 92108: contig of 2571 bp in length
92109 92208: gap of unknown length
92209 94902: contig of 2694 bp in length

```

*      94903      95002: gap of unknown length
*      95003      98082: contig of 3080 bp in length

Alignment Scores:
Pred. No.:      6,62e-30      Length:      193189
Score:          1311.00      Matches:      254
Percent Similarity: 94.91%      Conservative: 7
Best Local Similarity: 92.36%      Mismatches: 10
Query Match:    40.49%      Indels:      4
DB:             2      Gaps:      2

US-10-054-935-2 (1-614) x AC119462 (1-193189)

QY      1 MetThMetArgSerAlaValPheLySAIAlaAlaProAlaGlyLysAnProGlu 20
      |||
Db      160970 ATGACCATGATGCCAGCTCCAGGCGCGCCCTCCGCGCAACCCGAG 161029

QY      21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
      |||
Db      161030 CACGACTGGACTACAGCGGGCTGGCGGCGCCAGAGACGATCCGGGGCG 161089

QY      41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAla 60
      |||
Db      161090 GCCGAACCCACTCTCCCGCGGCATTCGTAAGCTCAAGAGCCCGCGCTGGCC 161149

QY      61 SerSerGlnGlyLysSerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79
      |||
Db      161150 TCTACCCAGGCGCGGAGCCCGCGCTCCAGCGCGCTGGCGCGGCAAGGCGCG 161209

QY      80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGluLeuSerTyrGlyLysSer 99
      |||
Db      161210 GGTTCGTACTCCCGCGCGGGGGCGCCCGCGGACAGAAAGAGAGTGGCGGTGCG 161269

QY      100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyLleGlyGlyGluProAla 119
      |||
Db      161270 GTCCCTTGCGCTGCTGGCGCCCGCGGCTACCAACAAAGCCGCGCATCGCGGGAGCCAGTC 161329

QY      120 AlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProIleGlnThr 139
      |||
Db      161330 GCAGCGCGCGCTGGCGAGCCCGCGCCCAATACAGCGCGTGCCTCCATTCAGACG 161389

QY      140 GlySerLeuValAlaAlaAlaLysGluProThrProThrAlaGlyLysAspLysGlyAla 159
      |||
Db      161390 GGTCTCTGCTGGCGCGCGGCAAGAGCCCTACGCTGGCTGGGAGCAAGGGTGGGGCG 161449

QY      160 AlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProProLeuProGly 179
      |||
Db      161450 GCTGCCCGAGCTGCCACCGGCTCGGACCGCGGGAGCCCGACACTACTCTGCGCGGG 161509

QY      180 ProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTyr 199
      |||
Db      161510 CCGCCACCCCTCGCGCGCCACCGCACCTGCGCGGAGCCCTGGCGCGGCAAGATGG 161569

QY      200 LysSerMetArgLysSerProLeuGlyGlyGlyGlySerLysAlaSerSerGlnAla 219
      |||
Db      161570 AAGAGATTAAGAGAGAGCCCTCTCGGGGGGGGGCGGCGCTGGGACCTTCATCGAGGCC 161629

QY      220 AlaCysLeuLysGlnLleLeuLeuGlnLeuAspLeuLleGluGlnGlnGlnGln 239
      |||
Db      161630 GCCTGGCTCAAAAGATTCCTTCGCTGCTGCAATTGGACTCATCGAAGCAGCAGCAGCAG 161689

QY      240 LeuGlnAlaLysGluLysGluLleGluGluLeuLysSerGluArgAspThrLeu----- 257
      |||
Db      161690 TTGCAAGCCCAAGAAAGAGATCGAGAGACTGAAATCGAGAGATACGATACGATACGGAG 161749

QY      258 ---LeuAlaArgLleGluArgMetGluArgArgMetGlnLeuVal 271
      |||
Db      161750 GGGTTAATCTGCATTAAGGATGGGAGGAGCGACTCTGCGCATGCTC 161794

RESULT 4
AX401049      615 bp      DNA      linear      PAT 06-JUN-2002
LOCUS
DEFINITION Sequence 725 from Patent WO0210453.

```

```

ACCESSION      AX401049
VERSION        AX401049.1      GI:21337229
KEYWORDS
SOURCE
ORGANISM
      Norway rat.
      Rattus norvegicus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
REFERENCE
AUTHORS      Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
      Blaschoff,M.R.
TITLE        Molecular toxicology modeling
JOURNAL      Patent: WO 0210453-A 725 07-FEB-2002;
      Gene Logic, Inc. (US)
FEATURES
      source
      location/Qualifiers
      1..615
      /organism="Rattus norvegicus"
      /db_xref="taxon:10116"
      /note="EMBL/Genbank Accession No. A1102093"
BASE COUNT      84 a      235 c      224 g      71 t      1 others
ORIGIN
US-10-054-935-2 (1-614) x AX401049 (1-615)

Alignment Scores:
Pred. No.:      1.99e-24      Length:      615
Score:          1055.50      Matches:      197
Percent Similarity: 97.55%      Conservative: 2
Best Local Similarity: 96.57%      Mismatches: 4
Query Match:    32.60%      Indels:      1
DB:             6      Gaps:      1

QY      13 AlaProAlaGlyLysAnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
      |||
Db      3 GCCCGCGCGCGGCAACCCGAGCGAGCGACTGATAGAGGGGCTGGCTCGGGCG 62

QY      33 GlyProGluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgLysLeu 52
      |||
Db      63 GGGCGCGAGAGCGATCGCGGGCGCGGCAAGCCCACTCTCCCGCGCATCGTAAGTC 122

QY      53 LysGluProGlyProProLeuAlaSerSerGlnGlySerProAlaProSerProAla 72
      |||
Db      123 AAGAGCCCGGGCGCGCGCTGCTTACCAAGGGAGGCCCGCGCTCTCCAGCT 182

QY      73 GlyCys---GlyLysGlyArgGlyLeuLeuProAlaGlyAlaAlaProGlyGln 91
      |||
Db      183 GGTGGGGGGCGGCAAGGGCGGGGTTGTACTCCGGCGGGCGCGCGCGGCGAG 242

QY      92 GlnGluGluSerTyrPheLysSerValProLeuProCysProProProAlaThrLysGln 111
      |||
Db      243 CAGGAAGAGAGCTGGGGGCTTCGTGCTTCCCTTCCTCGCGCCCGCGCTACCAACAA 302

QY      112 AlaGlyLleGlyLysGluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyr 131
      |||
Db      303 GCCGGCATCGGGGGAGCCAGTCGACGCGGGGCTGGCGAGCCCGCGGCAAGTAT 362

QY      132 GlnAlaValLeuProLleGlnThrGlySerLeuValAlaAlaLysGluProThrPro 151
      |||
Db      363 CAGGGGTCTGGCCATTCAGAGGGCTCTCTGCTGGCGGCGCAAGAGCTACGCC 422

QY      152 TyrPAlaGlyAspLysGlyLysAlaAlaSerProAlaAlaThrAlaSerAspProAlaGly 171
      |||
Db      423 TGGGCTGGGGAGACAGGGTGGGGCTCCCGCGCTGCGCACCGCTTCGAGACCGGGGGA 482

QY      172 ProProProLeuProLeuProGlyProProProProAlaProThrAlaThrAlaGlyThr 191
      |||
Db      483 CCGCCACCTACTCTTGGCGGGCGGCCACCCCTCGGCCACCGCATCGCGGAGC 542

QY      192 LeuAlaAlaSerGluLysArgTyrLysSerMetArgLysSerProLeuGlyLysGly 211
      |||
Db      543 CTGGCGGCGAGAGGAGGAGATGAAAGATTAAGAAAGAGCCCTCTCGGGGTGGCGGC 602

QY      212 GlySerGlyAla 215

```


DB 603 NGCCTCGGAGGCC 614

|||||||

RESULT 5 AC131356/c 174346 bp DNA linear HTG 21-AUG-2002

LOCUS Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS

DEFINITION *** 55 unordered pieces.

ACCESSION AC131356.1 GI:22380622

VERSION AC131356.1

KEYWORDS HTG: HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1. (bases 1 to 174346)

Muzny,D,Marle,M, Metzger,M, Lee, A, Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D., Anyalobechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,Y., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davis,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinn,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Flinley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Gara,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guenara,W., Gunaratne,P., Healand,W., Hamil,C., Hamilton,C., Hernandez,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Marhney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelameh,O., Okunodu,G., Olarunpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Polidexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Qutro,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Saverly,G., Scheer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartbeyan,A., Sisson,I., Sitter,C.D., Smajd,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,D., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (21-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web Site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: 600J

Center clone name: CH230-278A19

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 109854 bases at least Q40

Consensus quality: 11811 bases at least Q30

Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 55 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1	1547:	contig of 1547	bp in length	
*	1548	1647:	gap of unknown length	
*	1648	3064:	contig of 1417	bp in length
*	3065	3164:	gap of unknown length	
*	3165	4885:	contig of 1721	bp in length
*	4886	4985:	gap of unknown length	
*	4986	6000:	contig of 1015	bp in length
*	6001	6000:	gap of unknown length	
*	6101	7171:	contig of 1071	bp in length
*	7172	7271:	gap of unknown length	
*	7272	8794:	contig of 1523	bp in length
*	8795	8894:	gap of unknown length	
*	8895	10368:	contig of 1474	bp in length
*	10369	10468:	gap of unknown length	
*	10469	11906:	contig of 1438	bp in length
*	11907	12006:	gap of unknown length	
*	12007	13031:	contig of 1025	bp in length
*	13032	13131:	gap of unknown length	
*	13133	14575:	contig of 1444	bp in length
*	14576	14675:	gap of unknown length	
*	14676	15699:	contig of 1024	bp in length
*	15700	15799:	gap of unknown length	
*	15800	16951:	contig of 1152	bp in length
*	16952	17051:	gap of unknown length	
*	17052	18695:	contig of 1644	bp in length
*	18696	18795:	gap of unknown length	
*	18796	20470:	contig of 1675	bp in length
*	20471	20570:	gap of unknown length	
*	20571	22418:	contig of 1448	bp in length
*	22419	22518:	gap of unknown length	
*	22519	24364:	contig of 1846	bp in length
*	24365	24464:	gap of unknown length	
*	24465	25386:	contig of 1122	bp in length
*	25387	25686:	gap of unknown length	
*	25687	27859:	contig of 2173	bp in length
*	27860	27959:	gap of unknown length	
*	27960	29649:	contig of 1690	bp in length
*	29650	29749:	gap of unknown length	
*	29750	31331:	contig of 1382	bp in length
*	31332	31431:	gap of unknown length	
*	31432	33606:	contig of 2175	bp in length
*	33607	33706:	gap of unknown length	
*	33707	35179:	contig of 1473	bp in length
*	35180	35279:	gap of unknown length	
*	35280	37193:	contig of 1914	bp in length
*	37194	37293:	gap of unknown length	
*	37294	38382:	contig of 1089	bp in length


```

* 71983 72082: gap of unknown length
* 72083 74289: contig of 2207 bp in length
* 74289 74389: gap of unknown length
* 74390 76937: contig of 2548 bp in length
* 76938 77037: gap of unknown length
* 77038 79003: contig of 1966 bp in length
* 79004 79103: gap of unknown length
* 79104 81519: contig of 2416 bp in length
* 81520 81619: gap of unknown length
* 81620 84864: contig of 3245 bp in length
* 84865 84965: gap of unknown length
* 84966 87539: contig of 2575 bp in length
* 87540 87640: gap of unknown length
* 87641 89437: contig of 1798 bp in length
* 89438 89538: gap of unknown length
* 89539 92108: contig of 2571 bp in length
* 92109 92208: gap of unknown length
* 92209 94902: contig of 2694 bp in length
* 94903 95002: gap of unknown length
* 95003 98082: contig of 3080 bp in length

```

Alignment Scores:

```

Pred. No.: 7, 09e-18 Length: 193189
Score: 906.50 Matches: 221
Percent Similarity: 43.60% Conservative: 14
Best Local Similarity: 41.00% Mismatches: 55
Query Match: 28.00% Indels: 249
DB: 2 Gaps: 8

```

US-10-054-935-2 (1-614) x AC119462 (1-193189)

```

QY 149 ProThrProTPrAlaGlySlySlyGlyAlaAla-SerProAlaAlaThrLaserA 168
DB 186401 CCGGACAGGCGGAGATCATCTTCAGAGAGATGATCTP----- 186362
QY 168 ProAlaGlyProProleuProleuProleuProleuProleuProleuProleu 186
DB 186361 -CGGAGTGGTCCATCCCTATTACATCCAGGTTAACTTCATATTCATGCTTTT 186303
QY 186 rAlaThrAlaGlyThrLeuAlaAlaLaserGluGlyArgTrpLysSerMetArgLysSer 206
DB 186302 GCCTTACGCGGTTTAAAGGAACTATTCAGTGGTCCGTCCTCCATTCAGAGAAAGTTG 186243
QY 206 oleuGly-GlyGlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeuLysGln---- 224
DB 186242 AATTGGTAGGG-----TCAACAGCAACATGTAGAACGCCCAATT 186204
QY 225 -----IleLeuLeuGlnLeuAlaPLeuIleGluG 235
DB 186203 CCTACATGGGGTTAACTGCTGCTAATTCTATATGATTAAGCTGATCTGTGACTT 186144
QY 235 IncInGlnGlnLeuGlnAlaLysGluLysGluIleGluLysSerGluArgA 255
DB 186143 CC-----T 186141
QY 255 spThrLeuLeuAlaAglIleGluArgMetGluArgArgMetGlnLeuValLysLysAspA 275
DB 186140 TTATGCTCTGCTCGATTCGATTGAACGTATGAAAGCGGAGTGGTAAAGGAGTA 186081
QY 275 snGluLysGluArgHsLysLeuPheGlnGlyTrpGluTrpGluLysGluLysGln 295
DB 186080 ACAGAGAGAAAGGCAAGAGCTTTTCAGGGCTATGAACTGAGAGAGAGAGAGAAACGG 186021
QY 295 lueuSerGluLysIleLysLeuGluLysGlnProGluLeuSerGluThrSerGlnThrL 315
DB 186020 AGTTGCTGAGAAATTAACATGAGAGCGCAGCGGAGCTTCGAGAGACTCCAGAGCTC 185961
QY 315 eueProLysPProPheSerCysGlyArgSerGlyLysGlyHsLys----- 330
DB 185960 TGCCCTTAAGCCTTTCTCATGTGCGCGAGTGGCAAGGAGACAAAAGGTGTGTGTG 185901
QY 330 ----- 330

```

```

DB 185900 ACTTATTTCTATGAACTGAAAAAGATGAGAACTGCTAGTGTGANGAATATGCCCCTT 185841
QY 330 ----- 330
DB 185840 GATCTGTGATTTGGCTTTAGACACACCATAGCTGATTTGAGAAGTCTGTGTACTGGA 185781
QY 330 ----- 330
DB 185780 GCATGTGGCTTCCCTGTATATCAGACACACTAGTTCAGGGCCAGCCGTGAGTAGAA 185721
QY 330 ----- 330
DB 185720 CCTGTTTTTACTGTCTTGAGACATTTTTTTTGTACAGTCTTAAGTACAGAAATGCA 185661
QY 330 ----- 330
DB 185660 GAGTATATTTTAAAGAAATGGAATAGGGGTTGGGGATTTTACTGACGTGTAGACGCT 185601
QY 330 ----- 330
DB 185600 TGCCTAGCAAGCTCAAGGCCCTGGTTCGGTCCAGTCCGAGTCCGAAAAAAGAACCAAAAA 185541
QY 330 ----- 330
DB 185540 AAAAAAAGAAATAGTCTTAAGCATATAGACGCTTGTCTATGACCAAAAGGCTTTTCTGTA 185481
QY 330 ----- 330
DB 185480 TACTTTAAGCATGACACTTACTGTCATAGCTTCCAAATGAAACATGATATTTAGATA 185421
QY 330 ----- 330
DB 185420 CCTGGCTACCTCTTGTCCCTCAATATATTCAAATCAAAACTTATTTGAATTAATTA 185361
QY 331 -----ArgLysSerProPheGlySerThrGluArgL 341
DB 185360 TTTTATTAACTTTTCTTTCGTACCTAATATGAAAAACCCATTTGGAATATACGAAAGAA 185301
QY 341 ystThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysH 361
DB 185300 AGACTCTGTTAAAGCTGGGCTCTGAATTTTCAAAATTCAAAACAAAAAAGCTCTTAAGC 185241
QY 361 lserProIleLysGluGluProCysGlySerLeuSerGluThrValLysLysArgGluL 381
DB 185240 ACTCTCCATTAAGAGGAGGCCCTGTGTTCCATATCAGAAACTGTTGTAAACGTGAAT 185181
QY 381 eueArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLys 401
DB 185180 TGAAGAGCCAAAGAAACCCAGAAAAAGCCCGGCTTCATGTGATACCCACCAAGACTCT 185121
QY 401 erThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleG 421
DB 185120 CGACTCCCAAAAGGAGACCCAGACACCCAGCAAGGAGCAAGCTCTCAAGTAGATG 185061
QY 421 lueuSerProLysLeuSerThrThrGluMetLysLeuLysArgTrpHisGlnProProP 441
DB 185060 AAGATTTGCCGTACCTTTCCACACAGAAATGATTTGTCTGTGGCACACCTGCC 185001
QY 441 roserProleuProleuArgGluSerSerProLysLysGluLysGluThrValAlaArgCysL 461
DB 185000 CATCACCGTTTACCATTTACGGGAATCTCTCCAAAGAGAGAGAGACTGTAGCA----- 184948
QY 461 eueMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
DB 184947 -----AGTAAGCATATACGAACACTTCTCTT---ATACCTAGTGG 184909

```

```

RESULT 7
LOCUS HSM800235 1433 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFp586B1922 (from clone DKFp586B1922).
ACCESSION AL049450
VERSION AL049450.1 GI:4500236
KEYWORDS

```



```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14249
Center clone name: 54_L14
----- Summary Statistics
Sequencing vector: Plasmid; N/a; 100% of reads
Strategy: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 200414 bases at least Q40
Consensus quality: 203403 bases at least Q30
Insert size: 204607 bases at least Q20
Insert size: 210000; agarose-fp
Quality coverage: 7.0 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 9177: contig of 9177 bp in length
*
* 9178 9277: gap of 100 bp
*
* 9278 9931: contig of 654 bp in length
*
* 9932 10031: gap of 100 bp
*
* 10032 10809: contig of 778 bp in length
*
* 10810 10909: gap of 100 bp
*
* 10910 13046: contig of 2137 bp in length
*
* 13047 13146: gap of 100 bp
*
* 13147 14912: contig of 1766 bp in length
*
* 14913 15012: gap of 100 bp
*
* 15013 16787: contig of 1775 bp in length
*
* 16788 16887: gap of 100 bp
*
* 16888 19077: contig of 2140 bp in length
*
* 19078 19127: gap of 100 bp
*
* 19128 23490: contig of 4363 bp in length
*
* 23491 23590: gap of 100 bp
*
* 23591 25951: contig of 2361 bp in length
*
* 25952 26051: gap of 100 bp
*
* 26052 29248: contig of 3197 bp in length
*
* 29249 29348: gap of 100 bp
*
* 29349 32751: contig of 3403 bp in length
*
* 32752 32851: gap of 100 bp
*
* 32852 37464: contig of 4613 bp in length
*
* 37465 37564: gap of 100 bp
*
* 37565 42592: contig of 5028 bp in length
*
* 42593 42692: gap of 100 bp
*
* 42693 49347: contig of 6655 bp in length
*
* 49348 49447: gap of 100 bp
*
* 49448 54633: contig of 5186 bp in length
*
* 54634 54733: gap of 100 bp
*
* 54734 62208: contig of 7555 bp in length
*
* 62209 62388: gap of 100 bp
*
* 62389 146838: contig of 84450 bp in length
*
* 146839 146938: gap of 100 bp
*
* 146939 158635: contig of 11697 bp in length
*
* 158636 158735: gap of 100 bp
*
* 158736 175143: contig of 16408 bp in length
*
* 175144 175243: gap of 100 bp
*
* 175244 200086: contig of 24843 bp in length
*
* 200087 200186: gap of 100 bp
*
* 200187 207214: contig of 7028 bp in length.
*
Location/Qualifiers
1..207214

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-54U14"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .9177
/ote="assembly_fragment
clone_end:SP6
vector_side:left"
9278. .9931
/ote="assembly_fragment"
10032. .10809
/ote="assembly_fragment"
10910. .13046
/ote="assembly_fragment"
13147. .14912
/ote="assembly_fragment"
15013. .16787
/ote="assembly_fragment"
16888. .19027
/ote="assembly_fragment"
19128. .23490
/ote="assembly_fragment"
23591. .25951
/ote="assembly_fragment"
26052. .29248
/ote="assembly_fragment"
29349. .32751
/ote="assembly_fragment"
32852. .37464
/ote="assembly_fragment"
37565. .42592
/ote="assembly_fragment"
42693. .49347
/ote="assembly_fragment"
49448. .54633
/ote="assembly_fragment"
54734. .62288
/ote="assembly_fragment"
62389. .14638
/ote="assembly_fragment"
146939. .158635
/ote="assembly_fragment"
158736. .175143
/ote="assembly_fragment"
175244. .200086
/ote="assembly_fragment"
200187. .207214
/ote="assembly_fragment
clone_end:T7
vector_side:right"
39165 c 39030 g 61707 t 2001 others
BASE COUNT 65311 a
ORIGIN
Alignment Scores:
Pred. NO.: 0.00292 Length: 207214
Score: 416.00 Matches: 79
Percent Similarity: 90.53% Conservative: 9
Best local Similarity: 83.16% Mismatches: 9
Query Match: 12.85% Indels: 0
DB: 2 Gaps: 0
US-10-054-.935-2 (1-614) x AC107643 (1-207214)
Oy 305 GlnProgluIuSerGlnThrSerGlnThrIleuProProLySPropheSerCysGlyArg 324
Db 177968 AGGCCAGGGGATTTGGGAGACATCCCGAGGTTCGCTTCACAGCCCTTCTCATGTGGCCAG 178027
Oy 325 serGlyLysGlyHisIysArGlySerProphGlySerThrIrgIuArgLyThrProVal 344
Db 178028 AATGGAAAGGACACAAAGAAACCCATTTGAAATACAGAAAGAGACACCTCGTT 178087
Oy 345 LysLysIleuAlaProgluIuIuSerIysValIysThrIysThrProLySHisSerProIle 364

```

Db 178088 AAAAGCTGCTCCTGAAATTTTCAAAAGCTCAAAACAAAGCTTAAAGCTCTTACACTCTCCACATT 178147
 Oy 365 LysglugluProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGln 384
 |||||||.....
 Db 178148 AAAGAGCAACCTGTGGTTCATATCACAAACTGTTTAAACGGAATTGAGGACCA 178207
 Oy 385 GluThrProGluLysProArgSerSerValAspThrProArg 399
 |||||||.....
 Db 178208 GAAACCCCAAAAGATCGTCTTCACGAGGATACCCACCAAGA 178252

RESULT 9
 AC131356 174346 bp DNA linear HTG 21-AUG-2002
 LOCUS Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS
 DEFINITION *** 55 unordered pieces.
 AC131356
 AC131356.1 GI:22380622
 HTG: HTGS_PHASE1.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 174346)
 Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswalio,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E.,
 Cardenas,Y., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., pSouza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dlin,H., Divya,K.,
 Diaper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garra,M.,
 Gengeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guerrero,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
 Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Koyar,C.,
 Kovals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Maunhney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
 Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
 Morris,S., Muidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
 Newton,N., Nguyen,N., Norris,S., Nwackelmen,O., Okunonu,G.,
 Olarunpusogoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
 Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polidexter,A.,
 Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Qutroz,J., Rachlin,E.,
 Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
 Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
 Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S.,
 Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
 Sison,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
 Song,X.-Z., Sorelle,R., Sosa,T., Steimle,M., Strong,R., Sutton,A.,
 Sytek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
 Thigley,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
 Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
 Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlaczek,R.,
 Wooten,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
 Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
 Zhou,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
 Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 174346)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: G00J
 Center clone name: CH230-278A19
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 109854 bases at least Q40
 Consensus quality: 118111 bases at least Q30
 Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 55 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1547: contig of 1547 bp in length
 1548 1647: gap of unknown length
 1648 3064: contig of 1417 bp in length
 3065 4885: gap of unknown length
 4885 4885: contig of 1721 bp in length
 4886 6000: gap of unknown length
 6000 6000: contig of 1015 bp in length
 6001 6100: gap of unknown length
 6101 7171: contig of 1071 bp in length
 7171 7271: gap of unknown length
 7271 8794: contig of 1523 bp in length
 8795 8894: gap of unknown length
 8895 10368: contig of 1474 bp in length
 10369 10468: gap of unknown length
 10469 11906: contig of 1438 bp in length
 11907 12006: gap of unknown length
 12007 13031: contig of 1025 bp in length
 13032 13131: gap of unknown length
 13131 14575: contig of 1444 bp in length
 14576 14675: gap of unknown length
 14676 15699: contig of 1024 bp in length
 15700 15799: gap of unknown length
 15800 16951: contig of 1152 bp in length
 16952 17051: gap of unknown length
 17052 18695: contig of 1644 bp in length
 18696 18795: gap of unknown length
 18796 20470: contig of 1675 bp in length
 20471 22418: gap of unknown length
 22419 22518: contig of 1848 bp in length
 22519 24364: gap of unknown length
 24365 24464: gap of unknown length
 24465 25586: contig of 1122 bp in length
 25587 25686: gap of unknown length
 25687 27859: contig of 2173 bp in length
 27860 27959: gap of unknown length
 27960 29649: contig of 1690 bp in length
 29649 29749: gap of unknown length
 29750 31331: contig of 1582 bp in length
 31332 31431: gap of unknown length

TITLE

31122	33706	contig of 2175	bp	in length
33607	33706	gap of unknown	length	
33707	33179	contig of 1473	bp	in length
33580	33279	gap of unknown	length	
37194	37193	contig of 1914	bp	in length
37294	37193	gap of unknown	length	
38283	38382	contig of 1089	bp	in length
38483	38482	gap of unknown	length	
40760	40759	contig of 2277	bp	in length
40860	40859	gap of unknown	length	
42764	42763	contig of 1904	bp	in length
42864	42863	gap of unknown	length	
46000	45999	contig of 3136	bp	in length
46100	46099	gap of unknown	length	
49140	49139	contig of 3040	bp	in length
49240	49239	gap of unknown	length	
51763	51762	contig of 2523	bp	in length
51863	51862	gap of unknown	length	
54908	54907	contig of 3045	bp	in length
55007	55007	gap of unknown	length	
55124	55124	contig of 2207	bp	in length
57215	57314	gap of unknown	length	
57315	60299	contig of 2985	bp	in length
60300	60399	gap of unknown	length	
60400	62940	contig of 2541	bp	in length
62941	63040	gap of unknown	length	
63041	65074	contig of 2034	bp	in length
65075	65174	gap of unknown	length	
65175	68290	contig of 3116	bp	in length
68291	68390	gap of unknown	length	
68391	70496	contig of 2106	bp	in length
70497	70596	gap of unknown	length	
70597	73863	contig of 3267	bp	in length
73864	73963	gap of unknown	length	
73964	76872	contig of 2909	bp	in length
76873	76972	gap of unknown	length	
76973	79213	contig of 2243	bp	in length
79213	79315	gap of unknown	length	
79316	83513	contig of 4198	bp	in length
83514	83613	gap of unknown	length	
83614	87309	contig of 3696	bp	in length
87310	87409	gap of unknown	length	
87410	92353	contig of 4944	bp	in length
92354	92453	gap of unknown	length	
92454	954901	contig of 3448	bp	in length
95902	96001	gap of unknown	length	
96002	100934	contig of 4933	bp	in length
100935	101034	gap of unknown	length	
101035	104875	contig of 3841	bp	in length
104876	104975	gap of unknown	length	
104976	106470	contig of 4495	bp	in length
106471	109570	gap of unknown	length	
109571	117766	contig of 8106	bp	in length
117767	117776	gap of unknown	length	
124311	124312	contig of 6536	bp	in length
124313	124412	gap of unknown	length	
124413	126938	contig of 4526	bp	in length
126939	128038	gap of unknown	length	
128039	135086	contig of 6548	bp	in length
135087	135686	gap of unknown	length	
135687	143121	contig of 7435	bp	in length
143122	143221	gap of unknown	length	
143222	148127	contig of 5906	bp	in length
149128	148227	gap of unknown	length	
149228	155513	contig of 7286	bp	in length
155514	156513	gap of unknown	length	

Alignment Scores:	
Pred. No.:	0.0286
Score:	380.50
Percent Similarity:	27.62%
Best Local Similarity:	22.49%
Query Match:	11.75%
Length:	174346
Matches:	215
Conservative:	49
Mismatches:	183
Indels:	513

[illegible]

```

Db      57347 TTTGGGTAAGAC--GGAGGTTATTTATATATTT-----AC 57382
OY      294  |G|uleuSerGluIuylIleuGluGluGlnProGluLeuSerGluThrSerGlnTh 314
Db      57383 TAGGTTTATGTCATATATAGTATTTTGTGAACTGAGGTGAT----- 57428
OY      314  |LeuProProlysProPheSerCysGlyArgSerGlyIysGlyHisLysArgLysSerPr 334
Db      57428 ----- 57428
OY      334  |OphGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLysVa 354
Db      57429 -----GAGTTTGTCCTT 57442
OY      354  |LysThrLysThrProLysHisSerProIleLys--GluGluProCysGlySerLeuSe 373
Db      57443 GAAACGTGTACTCCAAATGCTGCTCATTTAGTACTGTCCCTGACCCCTCCAGAAACCTGA 57502
OY      373  |rGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSe 393
Db      57503 TGACAGTGTGTTTCAAGAGGACATGCAGAACTGAGCTAGACGAGAAAGAGAGAAAA-- 57560
OY      393  |rValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGl 413
Db      57561 -----CGGTAAAGGCCAAGACTCCCATCTCGGAG 57589
OY      413  |uLysAlaPheSerSerGluIleGluAspLeuProThrLeuSerThrThrGluMetLysLe 433
Db      57590 CATCTGAGCCCTAGAGACTGTGACTCTCTCCATAC----- 57626
OY      433  |uCysArgThrPheHisGlnPro-----ProPheSerProLeuProLeuArgGluSe 449
Db      57627 -----TGAGAGAGTCTGGGGCAGTGTGCTCATAGTCA----- 57659
OY      449  |rSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGl 469
Db      57660 -----GCTGAGTGTCTTAGCGGGGCGGCTGTCTCTTACA 57694
OY      469  |uThrSerValLeuAlaValProSerThrArgAsp----- 480
Db      57695 GGCACACATGTGGCCTTT--TCATGGATGTCCTCCCTCAGTACTACTTGTGTGGAAC 57751
OY      481  -----HisSerValGluProLeuArgAspProAsn-- 490
Db      57752 ACCAACACAAATCTAGAGTGTCTACACCCAGTGAETTCCTCT-----CAGCCAGGTAG 57805
OY      491  -----ProSerAspLeuGluAsnLeuAspSerValAlaPheSerLysArgHisAl 508
Db      57806 TCAGTTCCTTACCTGAGTAGAGGGTTAAACAAGACCTTCTAGCCTCGAAAGTG 57865
OY      508  |aLysLeuGluLeuAspGluLysArgArg----- 517
Db      57866 CAAACTGCTTTTGTGA-GAACAACTTCTAGTAAGAGTAATGCTTGTGTTGTTGTGT 57924
OY      518  -----LysArgThrAspIleGlnArgIleArgGluGlnArgIleLeuGlnAr 533
Db      57925 TTGCTGTGTTTTCACAGATGGGATATTCAGAGGATCAGAGGAAACAAAGATTTTTCAGCG 57984
OY      533  |gLeuGlnLeuArgMetLysLysLysLysGlyIleGlnGlnSerGluProGluValThrSe 553
Db      57985 ACTGACGCTCAGAAATGTATMAAAGAAAGAAATCTCAGAAATCTGAGCCTGAGGTTACCTC 58044
OY      553  |rPhePheProGluProAspAsp----- 560
Db      58045 ATTTTTCCTGAGCCGATGAGCGTAACTGTATCTATCTAATAAATGTAGAACTTTGGGG 58104
OY      560 ----- 560
Db      58105 GTTGGGATTTAGTCACTGCGAGAGCCCTGCGTTCACTCCCAAGCTCCGAAAAAAG 58164
OY      560 ----- 560

```

```

Db      58165 AAAAAAAAAATGTAGACTTTGGGGGTTGGTGAAGTGTGGCTCAGTATTAAGACATG 58224
OY      560 ----- 560
Db      58225 TCTGTTCTTGGACAGACCAAAATTTGTTTCCAGACGCCACATTGGCGACTCAGAACTG 58284
OY      560 ----- 560
Db      58285 CATGTACTCCAGGGAGCCCAATGCATTTTCTGACCTCGAAGGGCACTCAGTGCATCA 58344
OY      560 ----- 560
Db      58345 CATGTACAAGACACACAGACAGATTAACATATATATATAAATGAATGAATAG 58404
OY      560 ----- 560
Db      58405 AAATTTAAGTTGACAGACTTTTCAAGGGGACAAAAAGGTGAGGCTCCACAGCAAAAC 58464
OY      560 ----- 560
Db      58465 CTTTGTGATGAGTGGGCTGTGCTAAGAGAGTGTATTTTAAACATTTCTTACCTCT 58524
OY      561 -----ValGluSerLeuMetIleThrProPheL 570
Db      58525 GCTAACAGAGTCCCATGTGTTTCTTCCATATGAAGTTTGATGATACCCCTCTCT 58584
OY      570  |euProValAlaPheGlyArgProLeuProLysLeuThrProGln 585
Db      58585 TGCCGTGTTAGCATTTGTGAGCGGCATTAACAAATTAATGACACAGTAAGTAGAATGA 58644
OY      585 ----- 585
Db      58645 TTGTCCCTCTCTTTTCTCAGATAGTTTGTACCTCTTCAAGTCCATGGAAGTCAAGAT 58704
OY      585 ----- 585
Db      58705 AGAATGGGAGCCACACATGACTTAGGGAAGCATGAGCAAGCAAGATGAATTCCTT 58764
OY      585 ----- 585
Db      58765 TAATCAGAGTAGATGACTGTGTGCTACGTCCTACTTATGGCTTGTGAGTAGCCTTGTCACTT 58824
OY      586 -----AsnPheGluLeuProThrLeuAspGluArgSerArgCysArgL 600
Db      58825 CTTTTTTTTTGTCTAGAACTTTGAACTACCTGTTGTGATGAGGAGGAGCCGTTGCATAT 58884
OY      600  |euGluIleGlnLysLysGlnThrProHisArgThrCysArgLys 614
Db      58885 TGGAGATCCAGAAAGCAACACCTCACCGGACTTGTAGAGAG 58928
OY      614 ----- 614

```

RESULT 10
 AC123367
 LOCUS
 DEFINITION
 *** 39 unordered pieces.
 ACCESSION
 AC123367
 VERSION
 AC123367.2 GI:21903263
 KEYWORDS
 HTG; HTGS; PHASE1.
 SOURCE
 Rattus norvegicus.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 14577)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
 Alsbrooks,S.L., AmaralLunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbarella,D., Benton,J., Blinze,K., Blankenburg,K., Bonini,D.,
 Bouck,D., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,T., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homai,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwkw,S., Ogun,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivas,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Soderren,E., Sotake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleceyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 144577)
Worley,K.C.

Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 144577)
Worley,K.C.

Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:21240269.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center Project name: GWGC
Center Clone name: CH230-506C22

----- Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114009 bases at least Q40
Consensus quality: 118499 bases at least Q30
Consensus quality: 121325 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1699: contig of 1699 bp in length

1700 1799: gap of unknown length
1800 2876: contig of 1077 bp in length
2877 2976: gap of unknown length
2977 4389: contig of 1413 bp in length
4390 4489: gap of unknown length
4490 5708: contig of 1219 bp in length
5709 5808: gap of unknown length
5809 6967: contig of 1159 bp in length
6968 7067: gap of unknown length
7068 8623: contig of 1556 bp in length
8624 8723: gap of unknown length
8724 9993: contig of 1270 bp in length
9994 10093: gap of unknown length
10094 11908: contig of 1815 bp in length
11909 12009: gap of unknown length
12009 14386: contig of 2378 bp in length
14387 14865: gap of unknown length
14866 16346: contig of 1860 bp in length
16347 16447: gap of unknown length
16447 18733: contig of 2187 bp in length
18733 21914: contig of 3181 bp in length
21914 22014: gap of unknown length
22014 24433: contig of 2419 bp in length
24433 24533: gap of unknown length
24533 27931: contig of 3298 bp in length
27931 31839: gap of unknown length
31839 31939: contig of 3908 bp in length
31939 33680: gap of unknown length
33680 33780: contig of 1741 bp in length
33780 36869: gap of unknown length
36869 36870: contig of 3089 bp in length
36870 39041: gap of unknown length
39041 39141: contig of 2072 bp in length
39141 39142: gap of unknown length
39142 42086: contig of 2944 bp in length
42086 42185: gap of unknown length
42185 45561: contig of 3376 bp in length
45561 48095: gap of unknown length
48095 48195: contig of 2434 bp in length
48195 48196: gap of unknown length
48196 53241: contig of 4946 bp in length
53241 56744: gap of unknown length
56744 56745: contig of 3503 bp in length
56745 56845: gap of unknown length
56845 60010: contig of 3166 bp in length
60010 60110: gap of unknown length
60110 64117: contig of 4006 bp in length
64117 64217: gap of unknown length
64217 67814: contig of 3598 bp in length
67814 67914: gap of unknown length
67914 72543: contig of 4629 bp in length
72543 72544: gap of unknown length
72544 75853: contig of 3210 bp in length
75853 75953: gap of unknown length
75953 80517: contig of 4564 bp in length
80517 80617: gap of unknown length
80617 84561: contig of 3844 bp in length
84561 84562: gap of unknown length
84562 89191: contig of 4630 bp in length
89191 89291: gap of unknown length
89291 95800: contig of 6509 bp in length
95800 95900: gap of unknown length
95900 102912: contig of 7012 bp in length
102912 103012: gap of unknown length
103012 107519: contig of 4507 bp in length
107519 107619: gap of unknown length
107619 113616: contig of 5997 bp in length
113616 113716: gap of unknown length
113716 120950: contig of 7234 bp in length
120950 121050: gap of unknown length
121050 127758: contig of 6708 bp in length
127758 127858: gap of unknown length

* 127859 136315: contig of 8457 bp in length
 * 136316 136415: gap of unknown length
 * 136416 144577: contig of 8162 bp in length.
 Location/Qualifiers

FEATURES

1. 144577

/organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-506C22"

BASE COUNT 45381 a 25293 c 25234 g 42138 t 6531 others
 ORIGIN

Alignment Scores:

Pred. No.: 6.48 Length: 144577
 Score: 299.00 Matches: 80
 Percent Similarity: 71.09% Conservative: 11
 Best Local Similarity: 62.50% Mismatches: 22
 Query Match: 9.23% Indels: 18
 DB: 2 Gaps: 4

US-10-054-935-2 (1-614) x AC123367 (1-144577)

Oy 354 ValysThrylThrProlys---HisSerProIlelyGluGluProCysGlySerLeu 372
 Db 88046 GTCAAAACAAACCTCTACACATTAACTTACATTAAAGAGAACCTGTGGTTCATA 88105
 Oy 373 SerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysProArgSer 392
 Db 88106 TCAGAAACCGTTTGTAAACGAGANTGAGGGGCCAAGAAACCCAGAAACCCAGGTCT 88165
 Oy 393 SerValAspThrProProArgLeuSerThrProGluLysGlyProSerThrHisProlys 412
 Db 88166 TCAGTGTATACCCACCAAGACTCTGAGTCCCA----- 88201
 Oy 413 GluLysAlaPheSerSerGluIleGluAspLeuProtyrLeuSerThrThrluMetTyr 432
 Db 88202 ---AAGCT-CTCTCAGTGAAGATGAGAGT-TTGGCCGACCTGTCCACACGAAATGTAT 88256
 Oy 433 LeuCysArgTrpHisGlnProProProSerProLeuProLeuArgLysSerProlys 452
 Db 88257 TTATGTCTCTGGGACACCAATTCCTCCGATCACCATTACATTACGGGAGATCATCTCCA--- 88313
 Oy 453 LysGluGluThrValAla-----ArgCysLeuMetProSerSerValAlaGly 468
 Db 88314 AAGGAGGAGACGGTGTGCAAGTAAAGCATAGAGAACATTATACCCC--AGTGTGCCGCT 88371
 Oy 469 GluThrSerValLeuAlaValPro 476
 Db 88372 CAAGCTAATCTGTCTCTATGCT 88395

RESULT 11
 AC107643/C 207214 bp DNA linear HTG 21-APR-2002
 LOCUS Mus musculus clone RP23-54L14, WORKING DRAFT SEQUENCE, 21 ordered
 DEFINITION pieces.
 ACCESSION AC107643.2 GI:20258435
 VERSION AC107643.2
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 207214)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-54L14
 Unpublished
 2 (bases 1 to 207214)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhallier,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collinore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Laroque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
 MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
 McKernan,P., McKernan,K., Melidim,U., Menus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,P., Norbu,C.,
 Norman,K., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 207214)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boukhallier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collinore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Laroque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McKernan,P., McKernan,K., Melidim,U., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,P.,
 Norbu,C., Norman,K., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 21, 2002 this sequence version replaced g1:18308342.
 All repeats were identified using RepeatMasker.
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 54.L.14
 Center clone name: 54.L.14
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version: 0.960731
 Consensus quality: 200414 bases at least Q40
 Consensus quality: 203403 bases at least Q30
 Consensus quality: 204607 bases at least Q20
 Insert size: 21000; agarose-1p
 Insert size: 205214; sum-of-contigs
 Quality coverage: 7.0 in Q20 bases; agarose-1p
 Quality coverage: 7.1 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have

```
* provided by the submittor.
```

```
* This sequence will be replaced
```

```
* by the finished sequence as soon as it is available and
```

```
* the accession number will be preserved.
```

```
1          9177: contig of 9177 bp in length
```

```
    9178  9277: gap of      100 bp
```

```
    9278  9931: contig of 654 bp in length
```

```
    9932 10031: gap of      100 bp
```

```
    10032 10809: contig of 778 bp in length
```

```
    10810 10909: gap of      100 bp
```

```
    10910 13046: contig of 2137 bp in length
```

```
    13047 13146: gap of      100 bp
```

```
    13147 14912: contig of 1766 bp in length
```

```
    14913 15012: gap of      100 bp
```

```
    15013 16787: contig of 1775 bp in length
```

```
    16788 16887: gap of      100 bp
```

```
    16888 19027: contig of 2140 bp in length
```

```
    19028 19127: gap of      100 bp
```

```
    19128 23490: contig of 4363 bp in length
```

```
    23491 23590: gap of      100 bp
```

```
    23591 25951: contig of 2361 bp in length
```

```
    25952 26051: gap of      100 bp
```

```
    26052 29248: contig of 3197 bp in length
```

```
    29249 29348: gap of      100 bp
```

```
    29349 32751: contig of 3403 bp in length
```

```
    32752 32851: gap of      100 bp
```

```
    32852 37464: contig of 4613 bp in length
```

```
    37465 37564: gap of      100 bp
```

```
    37565 42592: contig of 5028 bp in length
```

```
    42593 42692: gap of      100 bp
```

```
    42693 49347: contig of 6655 bp in length
```

```
    49348 49447: gap of      100 bp
```

```
    49448 54633: contig of 5186 bp in length
```

```
    54634 54733: gap of      100 bp
```

```
    54734 62288: contig of 7555 bp in length
```

```
    62289 62388: gap of      100 bp
```

```
    62389 146838: contig of 84450 bp in length
```

```
    146839 146938: gap of      100 bp
```

```
    146939 158635: contig of 11697 bp in length
```

```
    158636 158735: gap of      100 bp
```

```
    158736 175143: contig of 16408 bp in length
```

```
    175144 175243: gap of      100 bp
```

```
    175244 200086: contig of 24843 bp in length
```

```
    200087 200186: gap of      100 bp
```

```
    200187 207214: contig of 7028 bp in length.
```

```
Location/Qualifiers
```

```
1..207214
```

```
/organism="Mus musculus"
```

```
/db_xref="taxon:10090"
```

```
/clone_1fb="RP23-54L14"
```

```
/clone_1fb="RPcT-23 Female Mouse BAC"
```

```
1..9177
```

```
/note="assembly_fragment
```

```
clone_end:SP6
```

```
vector_side:left"
```

```
9278..9931
```

```
/note="assembly_fragment"
```

```
10032..10809
```

```
/note="assembly_fragment"
```

```
10910..13046
```

```
/note="assembly_fragment"
```

```
13147..14912
```

```
/note="assembly_fragment"
```

```
15013..16787
```

```
/note="assembly_fragment"
```

```
16888..19027
```

```
/note="assembly_fragment"
```

```
19128..23490
```

```
/note="assembly_fragment"
```

```
23591..25951
```

```
/note="assembly_fragment"
```

```
26052..29248
```

```
/note="assembly_fragment"
```

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	REFERENCE
misc_feature			29349..32751								
misc_feature			/note="assembly_fragment"								
misc_feature			32852..37464								
misc_feature			/note="assembly_fragment"								
misc_feature			37565..42592								
misc_feature			/note="assembly_fragment"								
misc_feature			42693..49347								
misc_feature			/note="assembly_fragment"								
misc_feature			49448..54633								
misc_feature			/note="assembly_fragment"								
misc_feature			54734..62288								
misc_feature			/note="assembly_fragment"								
misc_feature			62389..146838								
misc_feature			/note="assembly_fragment"								
misc_feature			146939..158635								
misc_feature			/note="assembly_fragment"								
misc_feature			158736..175143								
misc_feature			/note="assembly_fragment"								
misc_feature			175244..200086								
misc_feature			/note="assembly_fragment"								
misc_feature			200187..207214								
misc_feature			/note="assembly_fragment"								
BASE COUNT	65311 a	39165 c	39030 g	61707 t	2001	others					
ORIGIN											
Alignment Scores:											
Pred. No.:	47.1	Length:	207214								
Score:	274.50	Matches:	56								
Percent Similarity:	80.56%	Conservative:	2								
Best Local Similarity:	77.78%	Mismatches:	11								
Query Match:	8.48%	Indels:	3								
DB:	2	Gaps:	1								
US-10-054-935-2 (1-614) x AC107643 (1-207214)											
OY	390	ProargSerSerValAspThrProProArgLeuSerThrProGlnIysGlyProSerThr	409								
Db	177976	CCCTCGCCTTCAGTGTATACCCACCACACACACTCTGCACACGCCCAAAAGAGCCACCTCC	177917								
OY	410	HisProIysGlnIysAlaIaphSerSerGluIleGlnAspLeuProIyrlLeuSerThrThr	429								
Db	177916	CATCCCAAGAGAAAGACCTTCGAACTGAGATGAGATTTGGCGTACTTTTCCATACCA	177857								
OY	430	GluMetIyrlLeuCyArGTrpHisGlnProProProSerProLeuProLeu-----	446								
Db	177856	GAAATGTAATTGTGTGTGGACAGACGCTCCCCATCAACGTTACTGTAGGGTTAGGG	177797								
OY	447	ArgGluSerSerProIysIysGlnGluThrValAla	458								
Db	177796	TTAGATCTCTCCCAAGAGAGGAGACTGTAGCA	177761								
RESULT 12											
LOCUS	CNS08CB6	136755 bp	DNA	linear	HTG 12-AUG-2002						
DEFINITION	Oryza sativa chromosome 12 clone OSJNB0108014, *** SEQUENCING IN										
ACCESSION	AL845344										
VERSION	AL845344.1	GI:22208185									
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.										
SOURCE	Oryza sativa.										

[illegible]

```

Db      1582  CGCTGGGCT-----ACCCGGCGCGCGCGCGGAGACCTTCACGGGG 1626
Qy      485  ProleuAryasProasPro 491
        ||| ||| |||
Db      1627  CCGCGCGCGCGCGCGCGCGG 1647

RESULT 14
AB020652                3385 bp  mRNA  linear  PRI 16-JUN-1999
LOCUS      Homo sapiens mRNA for KIA0845 protein, partial cds.
DEFINITION AB020652
ACCESSION  AB020652
VERSION     AB020652.1  GI:4240178
KEYWORDS
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:plu
            SK plus clone:hk05234.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (sites)
AUTHORS    Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE      Prediction of the coding sequences of unidentified human genes.
            XII. The complete sequences of 100 new cDNA clones from brain which
            code for large proteins in vitro
JOURNAL    DNA Res. 5 (6), 355-364 (1998)
MEDLINE    2 (bases 1 to 3385)
REFERENCE  Ohara,O., Suyama,M., Kikuno,R., Nagase,T. and Ishikawa,K.
AUTHORS    Direct Submission
TITLE      Submitted (02-DEC-1998) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdna@nfokazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES
SOURCE      Location/Qualifiers
            1..3385
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="hk05234"
            /sex="male"
            /issue_type="brain"
            /clone_lib="plu
            /dev_stage="adult"
            1..3385
            /gene="KIAA0845"
            <1..2802
            /gene="KIAA0845"
            /codon_start=1
            /product="KIAA0845 protein"
            /protein_id="BAA74868.1"
            /db_xref="GI:4240179"
            /translation="SRSEKEDQLALNDRLNPAFYIDKVNRLAHNRSLSGEAALNRQOA
            GRSMGLYERVEMRGAVYRIGCAARGOLRQEHNLLEDIAHYRBLRDLDEAROREOA
            EAAKALARTRQEAARVDLQKAQLQDECGYLRHNDHEVVELGQIDGSSAONA
            QMAEETEDALCKDVTLSALRELRQLEHNAVSTIQSEWEFRVLDRISEAKVNTDMA
            RSQAEETTERARQLQARTTELEALQSLKSDSLERORSELEDRHQADIASYQDAIQDLPA
            ELRTKMEMAAQLREYDQILNVKALDIEIAAYRKLLEGECRIGFGPIPSLEGGLP
            KIPVSTHIKMAQSEKTIKVYSEKSEKIVYIEOETQVTEECRIEVEEKEKEGKEG
            EGEDEFEARCGEETSPSPAEASPEKAKSPKKEAKSPAKAKSPAEKASPE
            VKSPERKASPAKESPPKASPEKASPEKASPAKVKSPERKASPAKESPAKASPP
            EKASPVKEAKESPAEKSPVKEKASPAEKSPKASPTKPEAKSPKASPEKASPP
            SPEKEAKSPERKASPVKAESKSPKASPVKAKASPEKASPVKEKASPEKASPP
            VKEAKASPEKASPVKEKASPVKEKASPVKEKASPVKEKASPEKAKTLDVKSPEAKTP
            AKKEARSPADKPEKASPVKEKASPVKEKASPVKEKASPVKEKASPVKEKASPVKEE
            KEPEOVKVPKPKKAEKAPATPKTEPEKSDSKPEAKPKAPKPKVEKKEKPEKPK
            KESVVEAKKEAEADKKVPTPEKAPKAVYKEDAKPKREKTEVAKKPPDDAKKPEPK
            PAEKKEAPKPKDTRERKAKPKPEKPKTEAKAKDDTKLSKPSKPAEKAEKSSSTD
            QKDSKPEKPAEKEDCAAKGK"
BASE COUNT      1051 a      869 c      1037 g      428 t
ORIGIN
Alignment Scores:
Pred. No.:      10.1      Length:      3385

```



```

Db 1086 AGCATTACAGAGTGGACGCTGAGTACAGAACCAAGTGGAGATGGCCGCCAGCT 1145
QY 244 ulysgluilegluleu----- 250
Db 1146 GCGAGATACCAAGACCTGCTCAATGTCAGATGGCTGTGATATAGATATAGCCGCTTA 1205
QY 250 ----- 250
Db 1206 CAGAAAATCTCGAAGGTGAAGAGTGTGATTGGCTTGGCCCAATCTCTTCGCT 1265
QY 250 ----- 250
Db 1266 TCCAGAAAGACTCCCAAAATTCCTCTGTCTCCTACATTAAGGTGAAGCCGAGA 1335
QY 251 -----LysSerGluArgAspThrLeuLeuAlaArgIleGluArgme 264
Db 1326 GAAGATCAAAAGTGTGAGAAAGTCTGAGAAAGAAAGTGTGATT-----GTGGAGAAACA 1379
QY 264 tgluatgargmetgluleuVal-----LysLysAspAsnGlu 277
Db 1380 GACAGAGAGACCCCAAGTGTGACTGAGAAAGTGAAGAGAGAAAGAGCCCAAGA 1439
QY 277 sgluatgHlLysLeuPheGlnGlyTrgluThrGluGluArgGlu----- 292
Db 1440 GAGAGAGGCGCAAGAGAGAGAGAGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGA 1499
QY 293 -----GluThrGluLeuSerGlu 300
Db 1500 AACAAAGTCTCCCAAGCAGAGAGAGGCTGATCCCAAGAGAGAGAGAGAGAGAGAG 1559
QY 300 elysleuglucysglu----- 305
Db 1560 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619
QY 305 ----- 305
Db 1620 CCCAGCCGAGTCAAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1679
QY 306 -----ProGluLeuSerGluThrSerGluThrLeuProGlu 319
Db 1680 ACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1739
QY 319 opHeserCysGlyArgSerGlyLysGlyHisArgLysSerProPheLysThrGlu 339
Db 1740 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1790
QY 339 uArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrPr 359
Db 1791 GCGCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1850
QY 359 oLysHisSerProLysGluLeu-----ProCysGlu 370
Db 1851 GCGCAAGTCCCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1910
QY 370 ySerLeuSerGluThrValCysLysArgGluLeuArgSerGlu----- 385
Db 1911 GCGCAAGTCTCAAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964
QY 386 -----ThrProGluLysProArgSer 393
Db 1965 TGAGAGAGCCCAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2024
QY 393 rVal-----AspThrProArgLysSerThrProGlnLysGlyProSerThr 409
Db 2025 AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2084
QY 409 rHisPro-----LysGluLysAlaPheSerSerGluIleGluAs 422
Db 2085 GTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2144
QY 422 pleuProTyrLeuSerThrThrGluMetTyrLeuGlyArgTrpHisGlnProProPhe 442
Db 2145 GTCCCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2183

```

```

QY 442 rProLeuProLeuArgGlu-----SerSerProLysLysGlu 454
Db 2184 CAAGTCCCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2243
QY 454 uGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAl 474
Db 2244 AGAAGCCCAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2303
QY 474 aValProSerTrpArgAspHisSerValGluProLeuArgAspProAsn-----Pr 491
Db 2304 GTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2363
QY 491 oSerAspLeuLeuGluLysLeuAspSerValPheSerLysArgHisAlaLysLeuGlu 511
Db 2364 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2423
QY 511 uLeuAspGluLysArgArg-----LysArgTrpAspIleGluArg 524
Db 2424 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2483
QY 524 g---IleArgGluGluArgIleLeuGluArgLeuGluLeuArgMetTyrLysLysGlu 543
Db 2484 CCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2543
QY 543 yIleGlnGluSerGluProGluValThrSerPhePheProGluProAspAsp-----Va 561
Db 2544 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2603
QY 561 lGluSerLeuMetIleThrProPheLeuProValVal-AlaPheGlyArgProLeuProL 581
Db 2604 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2663
QY 581 ySerLeuThrProGlnAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeu 601
Db 2664 CGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
QY 601 lueGlnLysLysGlnThrProHisArg 610
Db 2712 AGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740

```

Search completed: July 14, 2003, 21:40:43
 Job time : 5099 secs